# **Outlier-Robust Phase Retrieval in Nearly-Linear Time**

Anonymous Author(s) Affiliation Address email

## Abstract

1	Phase retrieval is a fundamental problem in signal processing, where the goal is
2	to recover a (complex-valued) signal from phaseless intensity measurements. It is
3	known that natural non-convex formulations of phase retrieval do not have spurious
4	local optima. However, the theoretical analyses of such landscape results often rely
5	on strong assumptions, such as the sampling vectors being Gaussian distributed.
6	In this paper, we propose and study the problem of outlier robust phase retrieval.
7	We seek to recover a vector $x \in \mathbb{R}^d$ from n intensity measurements $y_i = (a_i^\top x)^2$ ,
8	where the sampling vectors $a_i$ 's are initially i.i.d. Gaussian, but a small fraction of
9	the $(a_i, y_i)$ pairs are adversarially corrupted.
10	Our main result is a near-sample-optimal and nearly-linear-time algorithm that
11	provably recovers the ground-truth $x$ in the presence of adversarial corruptions.
12	We first solve a lightweight convex program to find a vector close to the ground
13	truth. We then run robust gradient descent starting from this initial solution,
14	leveraging recent advances in high-dimensional robust statistics. Our approach is
15	conceptually simple and provides a framework for developing robust algorithms
16	for other tractable non-convex problems.

## 17 **1 Introduction**

Phase retrieval is a fundamental problem in signal processing with applications in various fields, including electron microscopy [32], crystallography [33, 36], astronomy [11], and optical imaging [37].
In these applications, one often has access to only the magnitudes of the Fourier transforms of a
complex signal. This is because measuring magnitude (e.g., by aggregating energy over time) is
much easier than measuring phase (which requires detecting rapid changes). We refer the reader to
the survey articles [37, 26] for more details about the theory and applications of phase retrieval.

In this paper, we focus on the real-valued generalized phase retrieval problem, where the Fourier transform is replaced by a general linear operator. We first give a formal definition of this problem.

**Definition 1.1** (Phase Retrieval). Let  $x \in \mathbb{R}^d$  be the ground-truth vector. Let  $a_1 \dots a_n \in \mathbb{R}^d$  be nsampling vectors and let  $y_i = \langle a_i, x \rangle^2 \in \mathbb{R}$  be the corresponding intensity measurements. Given  $(a_i, y_i)_{i=1}^n$  as input, the task is to recover x.

Note that it is impossible to distinguish between x and -x, so it is sufficient to recover either one. 29 Under certain assumptions (e.g., when the  $a_i$ 's are Gaussian distributed), the phase retrieval problem 30 in Definition 1.1 can be solved in polynomial time with provable recovery guarantees. This was first 31 achieved via approaches based on semidefinite programming (SDP) relaxations (see, e.g., Candès 32 et al. [5]). In practice, the problem is often solved using first-order optimization algorithms such as 33 gradient descent. It is well-established that, although many natural formulations of phase retrieval 34 have nonconvex objectives, all local optima are globally optimal under certain assumptions [34, 3, 40]. 35 An example of such objective function is the following: 36

minimize 
$$f(z) = \sum_{i=1}^{n} (y_i - \langle a_i, z \rangle^2)^2$$
 subject to  $z \in \mathbb{R}^d$ .

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

<sup>38</sup> However, existing analyses of such landscape results often rely on strong assumptions, such as the

sampling vectors  $a_i$ 's are i.i.d. Gaussian. Our work is motivated by the following questions: Can we

<sup>40</sup> relax the assumptions used in proving landscape results in many tractable nonconvex problems? In the <sup>41</sup> context of phase retrieval, what happens if a small fraction of the  $(a_i, y_i)$ 's are changed adversarially?

42 We focus on the following strong contamination model (see, e.g., [13]).

**Definition 1.2** ( $\epsilon$ -Corruption). An algorithm first specifies the number of samples n, and n samples are drawn independently from some unknown distribution D. The adversary is allowed to replace up to  $\epsilon n$  samples with arbitrary points. The modified set of n samples is then given to the algorithm as

46 input. We say that a set of samples is  $\epsilon$ -corrupted if it is generated by the above process.

<sup>47</sup> Under the  $\epsilon$ -corruption model for high-dimensional data, a common goal is to design efficient <sup>48</sup> algorithms that can achieve dimension-independent error guarantees. Early work in robust statis-<sup>49</sup> tics [42, 23, 25] provided sample-efficient estimators for various tasks, but with runtimes exponential <sup>50</sup> in the dimension. A recent line of work, initiated by [13, 28], has developed computationally efficient <sup>51</sup> robust algorithms for many fundamental high-dimensional tasks. There has been significant progress <sup>52</sup> in the algorithmic aspects of robust high-dimensional statistics (see, e.g., [12]).

<sup>53</sup> We now formally define the main problem that we pose and study in this paper.

**Problem 1.3** (Outlier-Robust Phase Retrieval). Let  $\epsilon > 0$ . Let  $x \in \mathbb{R}^d$  be the ground-truth vector with  $\|x\|_2 = 1$ . First, *n* sampling vectors  $(a_i)_{i=1}^n$  are drawn i.i.d. from  $\mathcal{N}(0, I) \in \mathbb{R}^d$ . Let  $y_i = \langle a_i, x \rangle^2$ be the corresponding intensity measurements. Then, an adversary arbitrarily corrupts an  $\epsilon$ -fraction of the  $(a_i, y_i)$ 's. Finally, the corrupted  $(a_i, y_i)$ 's are given to the algorithm as input. The task is to find a vector  $z \in \mathbb{R}^d$  such that  $\min\{\|z - x\|_2, \|z + x\|_2\} \leq \Delta$  for some precision parameter  $\Delta > 0$ .

Note that we allow corruption in both the sampling vectors  $a_i \in \mathbb{R}^d$  and the intensity measurements  $y_i \in \mathbb{R}$ . We would like to answer the following algorithmic question:

61 Can we design a provably robust and near sample-optimal algorithm for the 62  $\epsilon$ -corrupted phase retrieval problem (Problem 1.3) that runs in nearly-linear time?

## 63 **1.1 Our Results and Contributions**

In this paper, we answer the above question affirmatively. We first state the main result of our paper.

<sup>65</sup> **Theorem 1.4** (Main, Informal). *Consider the outlier-robust phase retrieval problem (Problem 1.3).* 

66 Let  $\Delta > 0$ . Given an  $\epsilon$ -corrupted set of  $n = \widetilde{\Omega}(d \log^2(1/\Delta))$  samples, we can compute  $z \in \mathbb{R}^d$  in

time  $\widetilde{O}(nd)$  such that  $\min(||z - x||_2, ||z + x||_2) \le \Delta$  with probability at least 0.8.

<sup>68</sup> Our algorithm has near-optimal sample complexity, because even without corruption, recovering the <sup>69</sup> ground-truth vector x in general requires  $\Omega(d)$  samples because there are d degrees of freedom in x. <sup>70</sup> Moreover, our algorithm runs in time nearly-linear in the size of the input, and provably recovers <sup>71</sup> the ground-truth vector x with arbitrary precision  $\Delta$ . The formal version of Theorem 1.4 is stated as <sup>72</sup> Theorem 3.1 in Section 3.

<sup>73</sup> We remark that the success probability of Theorem 1.4 can be boosted to  $1 - \delta$  for any  $\delta > 0$  by <sup>74</sup> incurring an additional factor of  $T = O(\log(1/\delta))$  in the sample complexity and runtime. We can <sup>75</sup> randomly partition the input into T equal-sized disjoint sets and run our algorithm on each set to <sup>76</sup> obtain T solutions  $Z = \{z_1, \ldots, z_T\}$ . If we output a solution  $z^*$  that has the maximum number of <sup>77</sup> points in Z within distance  $2\Delta$ , we can show that  $r(z^*) \leq 3\Delta$  with probability at least  $1 - \delta$ .

Our main conceptual contribution is to propose and study the outlier-robust phase retrieval problem, where a small fraction of the input data is adversarially corrupted. Note that we allow arbitrary corruption in both the sampling vectors  $a_i \in \mathbb{R}^d$  and the intensity measurements  $y_i \in \mathbb{R}$ . The nonconvex optimization landscape of phase retrieval is well understood when the  $a_i$ 's are Gaussian distributed, but the adversarial robustness of such landscape results is largely unexplored.

83 Our main technical contributions include the design and analysis of a near sample-optimal and nearly-

<sup>84</sup> linear time algorithm that provably solves the phase retrieval problem in the presence of outliers.

<sup>85</sup> Our approach provides a conceptually simple two-step framework for developing outlier-robust

algorithms for tractable nonconvex problems that combines the robustness of spectral initialization

and the efficiency of the subsequent robust gradient descent.

<sup>1</sup>We write  $G \subseteq [n]$  for the (remaining) good samples, and  $B = [n] \setminus G$  for the corrupted samples.

#### 88 **1.2** Our Approach and Techniques

<sup>89</sup> When there are infinite samples and no corruption, the objective function f(z) can be simplified as

$$f(z) = \mathop{\mathbf{E}}_{a \sim \mathcal{N}(0, I_d)} \left[ (\langle a_i, z \rangle^2 - y_i)^2 \right] = 3 \|x\|_2^4 + 3 \|z\|_2^4 - 2 \|x\|_2^2 \|z\|_2^2 - 4 \langle x, z \rangle^2.$$
(1)

<sup>90</sup> Even though f(z) is nonconvex, we know that it has no spurious local optima [34, 3, 40].

Our approach follows the general structure of Candès et al. [3], which uses a two-step procedure.

<sup>92</sup> The first step uses spectral techniques to find an initial guess that is close enough to the ground truth.

<sup>93</sup> The second step applies gradient descent to converge to the final solution. However, both steps are

susceptible to adversarial corruption. We develop nearly-linear time and provably robust algorithms

<sup>95</sup> for both steps and combine them to get our main result.

96 Step 1: Robust Spectral Initialization. When there is no adversarial corruption, the empirical 97 second-moment matrix  $Y = (1/n) \sum_{i=1}^{n} y_i a_i a_i^{\top}$  has expectation  $\mathbf{E}[Y] = I + 2xx^{\top}$ , so its top 98 eigenvector is close to x. However, the adversary can arbitrarily change the top eigenvector.

To circumvent this issue, we assign a (nonnegative) weight  $w_i$  to each sample, and let  $Y_w$  denote the weighted intensity-based second-moment matrix  $Y_w = \sum_{i=1}^n w_i y_i a_i a_i^{\top}$ . Ideally, if the weights w are uniformly distributed on the remaining clean samples, the top eigenvector of  $Y_w$  will align with x. We propose a novel optimization problem that can be used to find a weighting w such that  $Y_w$  must be close to the unknown unbiased expectation  $I + 2xx^{\top}$ . Moreover, we show that such a weight wcan be computed in nearly-linear time.

105 **Step 2: Approximate Gradient Descent.** Starting with the initial guess  $z_1 \in \mathbb{R}^d$  produced by the 106 robust spectral initialization, we want to apply gradient descent to recover the ground truth  $x \in \mathbb{R}^d$ . 107 Without corruption, if the initialization is close enough to x, each iteration will bring z closer to x by 108 a constant factor. This convergence guarantee can be compromised by the corrupted samples.

At a high level, approximating the gradient at a specific point amounts to a robust mean estimation problem (for the underlying distribution of the gradients). When the input data is  $\epsilon$ -corrupted, the gradients of the *n* samples can be viewed as an  $\epsilon$ -corrupted set of vectors. We can approximate the true gradient using this  $\epsilon$ -corrupted set of *n* gradients using robust mean estimation algorithms.

## 113 1.3 Related and Prior Works

Phase Retrieval. The problem of phase retrieval arises in many areas of science and engineering
[11, 33]. Early research on this problem proposes error-reduction algorithms [22, 17, 18]. Convex
and nonconvex optimization with various objective functions were later proposed and achieved exact
recovery [43, 3–5, 38]. Follow-up works generalize to robust phase retrieval where the observations
are subject to perturbations [45, 27, 7, 6, 31].

Nonconvex Optimization. Even though optimizing a nonconvex function is NP-Hard in general, 119 recent works showed that many nonconvex functions are locally optimizable due to discrete or 120 rotational symmetry. Besides phase retrieval, it is known that all local optima are globally optimal 121 for natural nonconvex formulations of a wide range of machine learning problems, such as matrix 122 completion [21], matrix sensing [2], phase synchronization [1], dictionary learning [39], and tensor 123 decomposition [20] (see also Chapter 7 of [44]). Closely related to our work, a recent line of work 124 explored the robustness of these landscape results: [30] studied matrix sensing in the  $\epsilon$ -corrupted 125 126 model and [8, 19] studied matrix completion and matrix sensing in semi-random models.

High-Dimensional Robust Statistics. Recent works in high-dimensional robust statistics developed
nearly-linear time algorithms for the problem of robust mean estimation [9, 16, 29]. Prior works [35,
14] developed meta-algorithms for finding *first-order* stationary points with dimension-independent
accuracy guarantees, which is closely related to the robust gradient descent procedure that we use.

#### 131 1.4 Roadmap

We first introduce notations and background in Section 2. Then we give an overview of our approach in Section 3. Next, we focus on how to get an initialization that is close enough to the ground truth xin Section 4. After the initialization, we use robust mean algorithms to estimate gradients to converge to the desired accuracy in Section 5. Finally, we conclude in Section 6 and discuss open problems.

## **136 2 Preliminary and Background**

**Notation.** We write [n] for the set of integers  $\{1, \ldots, n\}$ . We use  $\{e_1, \ldots, e_d\}$  for the standard unit 137 vector basis in  $\mathbb{R}^d$  and I for the identity matrix. For a vector x, we denote its  $\ell_1$ ,  $\ell_2$  and  $\ell_\infty$  norm as 138  $||x||_1, ||x||_2$  and  $||x||_{\infty}$ , respectively, and write the *i*<sup>th</sup> coordinate in x as  $x_i$ . For vectors  $x, y \in \mathbb{R}^d$ , 139 we denote its inner product as  $\langle x, y \rangle = x^{\top} y$ . For a matrix A, we use  $||A||_2$ ,  $||A||_*$ , and  $||A||_F$  as its 140 operator norm, nuclear norm, and Frobenius norm, respectively. We write  $\lambda_k(A)$  as the kth-largest 141 eigenvalues of A, and  $\overline{\lambda}_k(A)$  as the sum of the k largest eigenvalues. A symmetric  $n \times n$  matrix A 142 is said to be positive semidefinite (PSD) if for all vectors  $x \in \mathbb{R}^n$ ,  $x^{\top}Ax > 0$ . For two symmetric 143 matrices A and B, we write  $A \preceq B$  when B - A is positive semidefinite. 144

<sup>145</sup> **Packing SDP.** We will use nearly-linear time solvers for the following packing SDP.

$$\max_{w} \|w\|_{1} \qquad \text{subject to} \qquad \sum_{i=1}^{n} w_{i}A_{i} \preceq I, \quad \overline{\lambda}_{k} \left(\sum_{i=1}^{n} w_{i}B_{i}\right) \leq k, \quad w_{i} \geq 0, \forall i. \qquad (*)$$

**Lemma 2.1** ([10]). Given an instance of optimization (\*) with semi-positive definite matrices  $A_i \in \mathbb{R}^{d_1 \times d_1}$  and  $B_i \in \mathbb{R}^{d_2 \times d_2}$  with  $A_i = C_i C_i^\top$ ,  $B_i = D_i D_i^\top$  for all  $i = 1, 2, \dots, m$ , together with integer k > 0, error tolerance  $\epsilon_0 \ge 1/m^2$ , and failure probability  $\delta_0$ , there is an algorithm that runs in time  $\widetilde{O}((t_C + t_D + d_1 + d_2) \operatorname{poly}(1/\epsilon_0, \log 1/\delta_0))$ , where  $t_{C_i}$  and  $t_{D_i}$  are the time take to perform a matrix product with  $C_i$  and  $D_i$  respectively and  $t_C = \sum_{i=1}^n t_{C_i}$  and  $t_D = \sum_{i=1}^n t_{D_i}$ , and outputs w' with  $||w'||_1 \ge (1 - \epsilon_0)$ OPT where OPT is optimal value, with probability at least  $1 - \delta_0$ .

152 **Computing the Top Eigenvector.** We use power method to compute the top eigenvector of a matrix.

**Lemma 2.2** (Power Method for Top Eigenvector, e.g., [41]). Let  $A \in \mathbb{R}^{d \times d}$  and let  $\lambda_1$  be its largest eigenvalue. For any  $\overline{\delta} \in (0, 1)$ , there exists an algorithm that takes A and outputs a unit vector  $x \in \mathbb{R}^d$  in time  $O(t \log(d)/\overline{\delta})$  such that  $x^T A x \ge (1 - \overline{\delta})\lambda_1$  with probability at least 0.99, where t is the time required to compute Av for an arbitrary  $v \in \mathbb{R}^d$ .

**Robust Mean Estimation.** Another tool we use is robust mean estimation in the  $\epsilon$ -corruption model for distributions with bounded covariance. We use robust mean estimation algorithms to approximate the true gradient under adversarial corruption.

**Lemma 2.3** (Robust Mean Estimation, e.g., [15]). Let  $\mathcal{D}$  be a distribution on  $\mathbb{R}^d$  with unknown mean  $\mu$  and unknown covariance matrix  $\Sigma$  where  $\Sigma \leq \sigma^2 I$ . Let  $\epsilon_0$  be a sufficiently small universal constant. Let  $0 < \epsilon < \epsilon_0$  and  $\delta > 0$ . Given an  $\epsilon$ -corrupted set of n samples drawn from  $\mathcal{D}$ , we can output a vector  $\hat{\mu} \in \mathbb{R}^d$  in time  $\widetilde{O}(nd \log(1/\delta))$  such that, with probability at least  $1 - \delta - \exp(-n\epsilon)$ , we have  $\|\hat{\mu} - \mu\|_2 = O\left(\sqrt{\epsilon} + \sqrt{\frac{d}{n\delta}} + \sqrt{\frac{d(\log d + \log 1/\delta)}{n}}\right) \sigma$ .

$$\|\mu - \mu\|_{2} = O\left(\sqrt{c} + \sqrt{n\delta} + \sqrt{n}\right)$$

## 165 **3** Overview

166 We first state a formal version of our main result.

**Theorem 3.1** (Main). Consider the setting of Problem 1.3. Let  $0 < \epsilon < \epsilon'$  for some universal constant  $\epsilon'$  and let  $\Delta > 0$ . Given an  $\epsilon$ -corrupted set of  $n = \tilde{\Omega}(d\log^2(1/\Delta))$  samples, we can compute a vector  $z \in \mathbb{R}^d$  in time  $\tilde{O}(nd\log(1/\Delta))$  such that  $r(z) = \min\{||z - x||_2, ||z + x||_2\} \le \Delta$ with probability at least 0.8.

Theorem 3.1 requires two key technical lemmas: the robust spectral initialization (Lemma 3.2) and the approximate gradient descent (Lemma 3.3).

We first show that the spectral initialization can be done in nearly linear time with high probability,the proof of which can be found in Section 4.

**Lemma 3.2** (Robust Spectral Initialization). Under the setting of Problem 1.3, for any  $0 < \epsilon < \epsilon'$  for some universal constant  $\epsilon' > 0$ , given an  $\epsilon$ -corrupted set of  $n = \widetilde{\Omega}(d)$  samples, we can compute a vector  $z_0 \in \mathbb{R}^d$  of the ground truth x in time  $\widetilde{O}(nd)$  such that  $r(z_1) = \min\{||z_1 - x||_2, ||z_1 + x||_2\} \le \frac{1}{8}$ with probability at least 0.95. Then, with such initialization results, we can proceed to show that an approximate gradient descent algorithm can be used to find an arbitrary approximation of the ground truth in Section 5.

**Lemma 3.3** (Robust Gradient Descent). Consider the setting of Problem 1.3. Let  $\Delta > 0$  be the desired precision. Let  $0 < \epsilon < \epsilon_0$  for some sufficiently small universal constant  $\epsilon_0$ . Given an  $\epsilon$ -corrupted set of  $n = \tilde{\Omega}(d\log^2(1/\Delta))$  samples and an initial guess  $z_1$  such that  $r(z_1) = \min(||z_1 - x||_2, ||z_1 + x||_2) \le 1/8$ , we can compute a vector  $z \in \mathbb{R}^d$  in time  $\tilde{O}(nd)$  such that  $r(z) \le \Delta$  with probability at least 0.95.

For technical reasons, we cannot use the same set of samples for both the robust spectral initialization and the approximate gradient descent. Therefore, we partition the  $\epsilon$ -corrupted set of 2n samples into two equally sized disjoint sets, using one set for each algorithm.

*Proof of Theorem 3.1.* Let  $2n = \tilde{\Omega}(d \log^2(1/\Delta))$  be a set of  $\epsilon/2$ -corrupted samples. We partition the input into two disjoint sets of n samples. Both sets are  $\epsilon$ -corrupted. By Lemmas 3.2 and 3.3, for any  $\epsilon \in [0, \epsilon']$  and  $\Delta > 0$ , our algorithm takes the first set of samples and output a vector z' in time  $\tilde{O}(nd)$  such that  $r(z') \leq 1/8$  with probability at least 0.95. Then, using z' and the second set of samples, our algorithm can output  $z \in \mathbb{R}^d$  in time  $\tilde{O}(nd)$  such that  $r(z) \leq \Delta$  with probability at least 0.95. The overall success probability is at least 0.8, and the combined running time is  $\tilde{O}(nd)$ .

## **195 4 Robust Spectral Initialization**

We dedicate this section to proving Lemma 3.2: Given an  $\epsilon$ -corrupted set of  $(a_i, y_i)$ 's, we can compute an initial guess  $z_1 \in \mathbb{R}^d$  that is close to the ground truth x, where  $\min(||z_1 - x||_2, ||z_1 + x||_2) \leq 1/8$ . To build some intuition, consider the following intensity-based covariance matrix  $Y = \frac{1}{n} \sum_{i=1}^{n} y_i a_i a_i^{\top}$ , where each  $a_i$  is drawn independently from  $\mathcal{N}(0, I)$  and  $y_i = \langle a_i, x \rangle^2$ . The expectation of this matrix is  $\mathbb{E}[Y] = I + 2xx^{\top}$ . In other words, when there are enough samples and no adversarial corruption, we can obtain a good guess of the ground truth x (or -x) by computing the top eigenvector of Y. However, we cannot rely on this approach in adversarial settings.

To tackle this issue, we propose a nearly-linear time preprocessing step (Algorithm 1) that can recover the true expectation of Y under adversarial corruptions. Algorithm 1 assigns a non-negative weight to each sample. For a weight vector  $w \in \mathbb{R}^n$  and a set of indices  $S \subseteq [n]$ , the weighted intensity-based covariance matrix is defined as  $Y_{S,w} \doteq \sum_{i \in S} w_i y_i a_i a_i^\top$ , and we omit S when S = [n]. The feasible region for the weight vector is:  $\Delta_{n,\epsilon} := \left\{ w \in \mathbb{R}^n : \|w\|_1 = 1 \text{ and } \forall i \in [n], 0 \le w_i \le \frac{1}{(1-\epsilon)n} \right\}$ .

A weight w defines an empirical distributions over the samples  $(a_i, y_i)_{i=1}^n$ , where the largest probability assigned to any point is  $\frac{1}{(1-\epsilon)n}$ . Ideally, we would like to find a weight vector  $w^* \in \Delta_{n,\epsilon}$ that assigns its weight uniformly to all the uncorrupted samples, i.e.,  $w_i^* = \frac{1}{(1-\epsilon)n} \cdot \mathbb{1}_{i \in G}$ . To find a suitable weighting w, we use the following optimization problem (\*\*) in which  $\overline{\lambda}_2$  returns the sum of the top two eigenvalues (commonly known as the Ky Fan k norm for k = 2).

$$\min_{w} \quad \overline{\lambda}_2 \left( \sum_{i=1}^n w_i y_i a_i a_i^\top \right) \qquad \text{subject to} \quad 0 \le w_i \le \frac{1}{(1-\epsilon)n}, \forall i \in [n], \quad \sum_{i=1}^n w_i = 1. \quad (**)$$

At a high level, our main observation is that  $y_i a_i a_i^{\top}$  is always a positive semidefinite matrix as  $y_i \ge 0$ . Consequently, the adversary can only *add* extra directions with large eigenvalues, but will not be able to remove the eigendirection of x. By minimizing the Ky Fan 2 norm, we can remove any directions added by the adversary and make sure that the only remaining large eigendirection is close to x.

Let  $\delta \ge 0$  be some constant to be determined. We show that we can obtain a robust spectral initialization by solving the packing SDP problem (\*), which can be solved efficiently using Lemma 2.1. In particular, to fit the reweighting problem of (\*\*) into the framework of the generalized packing problem (\*), we define the following constraint matrices for all  $i \in [n]$ :

$$A_i := (1 - \epsilon)n \cdot e_i e_i^\top, \qquad B_i := \frac{1}{2}(1 - \delta)y_i a_i a_i^\top.$$

$$\tag{2}$$

The matrices  $(A_i)_{i=1}^n$  are used to implement the constraint that  $w \in \Delta_{n,\epsilon}$ . The matrices  $(B_i)_{i=1}^n$  help make sure the sum of the top two eigenvalues of  $Y_w$  must be at most roughly 4, because  $\overline{\lambda}_2(Y_{w^*}) \approx 4$ .

#### Algorithm 1 Robust Spectral Initialization

**Input:**  $\epsilon$ -corrupted samples  $(a_i, y_i)_{i \in [n]}$ 

**Output:** The initial guess  $z' \in \mathbb{R}^d$ 

1: function ROBUSTINIT( $\{(a_i, y_i)\}_{i \in [n]}$ )

- 2:  $\{A_i, B_i\} \leftarrow \text{Constraint matrices as defined in Equation (2)}$
- 3:  $w'' \leftarrow$  Solution to Optimization (\*) with constraints  $\{A_i, B_i\}, k = 2$ , and precision  $\epsilon_0 = 0.9$  as in Lemma 2.1
- 4:  $w' \leftarrow w'' / \|w''\|_1$
- 5:  $z_1 \leftarrow \text{TOPEIGENVECTOR}(Y_{w'})$  as in Lemma 2.2 with sufficiently small constant  $\overline{\delta}$ .
- 6: return  $z_1$
- 7: end function

First, we show that the weight w' computed by Algorithm 1 can ensure the weighted intensity-based covariance matrix  $Y_{w'}$  is close enough to the unbiased expectation  $I + 2xx^{\top}$ .

**Lemma 4.1.** With probability at least 0.98, the w' outputted by Algorithm 1 satisfies:

$$\|Y_{w'} - (I + 2xx^{\top})\|_{2} = O(\delta)$$
(3)

- In order to show Lemma 4.1, we need the following auxiliary Lemma 4.2, the proof of which can be found in Section A. Intuitively, Lemma 4.2 suggests that any weight w in the feasible region  $\Delta_{n,2\epsilon}$ will not have a huge impact on the properties of uncorrupted measurements.
- **Lemma 4.2.** For any  $\delta_0 > 0$ , and sufficiently small  $\epsilon \ge 0$ , given a set of  $n \epsilon$ -corrupted samples with
- 230  $n > \widetilde{\Omega}(d)$ , with probability at least 0.98, we have  $\|Y_{G,w} (I + 2xx^{\top})\|_2 \le \delta_0$  for all  $w \in \Delta_{n,2\epsilon}$ .
- Using Lemma 4.2, we provide a proof sketch for Lemma 4.1, and defer the details to Section A.
- *Proof.* We condition on the fact that the event of Lemma 4.2 holds (with probability at least 0.98) for  $\delta_0 = \delta$ . Thus, for the remaining of the proof, we assume that for all  $w \in \Delta_{n,2\epsilon}$ , it holds that  $\|Y_{G,w} - (I + 2xx^{\top})\|_2 \leq \delta$ .
- Let  $\lambda_1$  and  $\lambda_2$  be the top two eigenvalues of  $Y_{w'}$ , with  $v_1$  and  $v_2$  to be their corresponding eigenvectors.
- Note that the largest eigenvalue of  $I + 2xx^{\top}$  is 3, and the rest of the eigenvalues are all 1. In the proof, we show that the eigenvalues of  $Y_{w'}$  are also close to the ones of  $I + 2xx^{\top}$ . Our proof consists of two parts. We first establish lower bounds for  $\lambda_1$  and  $\lambda_2$ , and then find an upper bound for  $\lambda_1 + \lambda_2$ .
- **Lower Bound.** Since  $y_i a_i a_i^{\top} \succeq 0$  for any  $i \in [n]$ , for any positive weight vector  $w \in \Delta_{n,\epsilon}$ , we have  $Y_{G,w} \preceq Y_w$ . Thus a lower bound on eigenvalues of  $Y_{G,w'}$  will also be a lower bound on  $Y_{w'}$ .
- For the top eigenvalue  $\lambda_1$  of  $Y_{w'}$ , it holds

$$\lambda_1 = v_1^\top Y_{w'} v_1 \ge x^\top Y_{w'} x \ge x^\top Y_{G,w'} x \ge x^\top (I + 2xx^\top) x - \delta = 3 - \delta.$$
(4)

Similarly, for the second largest eigenvalue  $\lambda_2$  of  $Y_{w'}$ , we have:

$$\lambda_2 = v_2^\top Y_{w'} v_2 \ge v_2^\top Y_{G,w'} v_2 \ge v_2^\top (I + 2xx^\top) v_2 - \delta = 1 + 2 \langle v_2, x \rangle^2 - \delta \ge 1 - \delta.$$
 (5)

**Upper Bound.** Through the optimization problem (\*), a weight w'' is calculated such that  $Y_{w''}$ are operator-norm upper-bounded by the constraint parameters. Let OPT be the value of the optimal solution of the optimization problem (\*). The desired uniform weight vector over the good samples  $w^* \in \Delta_{n,2\epsilon}$  is also a feasible solution to this optimization problem because  $Y_{w^*}$  satisfy the optimization constraints due to Lemma 4.2. Since w'' is an  $\epsilon_0$ -approximation to the problem, we have  $\|w''\|_1 \ge (1 - \epsilon_0) \mathsf{OPT} \ge (1 - \epsilon_0) \|w^*\|_1 = 1 - \epsilon_0$ 

By optimization constraints, the Ky Fan 2-norm of  $\sum_i w_i'' B_i = \frac{1}{2}(1-\delta)Y_{w''} \le 2$ , and consequently,

$$\lambda_1 + \lambda_2 = \frac{1}{\|w''\|_1} \overline{\lambda}_2(Y_{w''}) \le \frac{4}{(1 - \epsilon_0)(1 - \delta)} \quad .$$
(6)

By combining inequalities (4), (5), and (6), we have shown that the top two eigenvalues of  $Y_{w'}$  are close to 3 and 1. Since the rest of the eigenvalues of  $Y_{w'}$  can also be bounded, we can conclude that  $\|Y_{w'} - (I + 2xx^{\top})\|_2 = O(\delta).$ 

- We can now show the closeness between the top eigenvector of  $Y_{w'}$  and the ground truth.
- **Lemma 4.3.** There exists an universal constant  $\epsilon'$  such that if  $0 \le \epsilon \le \epsilon'$ , and Algorithm 1 receives in input an  $\epsilon$ -corrupted set of samples, then it outputs  $z_1 \in \mathbb{R}^d$  such that with probability at least 0.95
- 255 *it holds*  $r(z_1) \le \frac{1}{8}$ .
- 256 Proof. We condition on the fact that the event of Lemma 4.1 holds (with probability at least 0.98).
- Let the eigendecomposition of  $Y_{w'}$  be  $Y_{w'} = \sum_{i \in [d]} \lambda_i v_i v_i^{\top}$ , where  $\lambda_1 \ge \ldots \ge \lambda_d$ . Under the basis  $\{v_1, \cdots, v_d\}$ , the ground truth x can be represented as  $x = \sum_{i \in [d]} \alpha_i v_i$ . Note that  $\|x\|_2^2 = \sum_{i \in [d]} \alpha_i^2 = 1$ . By Lemma 4.1, we have  $\|Y_{w'} - (I + 2xx^{\top})\|_2 = O(\delta)$ . Thus, we have

$$\begin{split} & x^{\top}Y_{w'}x \geq 3 - O(\delta) \quad \text{ and } \\ & x^{\top}Y_{w'}x \leq \lambda_1\alpha_1^2 + \lambda_2(1-\alpha_1^2) \leq (3+O(\delta))\alpha_1^2 + (1+O(\delta))(1-\alpha_1^2) \leq 1+2\alpha^2 + O(\delta). \end{split}$$

260 This implies  $\alpha_1^2 \ge 1 - O(\delta)$ . As a result,

$$r^{2}(v_{1}) = \left(\min\{\|v_{1} - x\|_{2}^{2}, \|v_{1} + x\|_{2}^{2}\}\right) = \min\{(\alpha_{1} - 1)^{2}, (\alpha_{1} + 1)^{2}\} + \sum_{i=2}^{a} \alpha_{i}^{2}$$
$$= \min\{2 - 2\alpha_{1}, 2 + 2\alpha_{1}\} = O(\delta).$$

The last inequality holds as long as  $\delta$  is sufficiently small. Let  $z_1 = \sum_{i \in [d]} \beta_i v_i$  be the unit vector approximating  $v_1$  returned by the algorithm. By Lemma 2.2, we have that  $z_1^\top Y_{w'} z_1 \ge (1 - \overline{\delta})\lambda_1$ with probability at least 0.99. Thus, we have:

$$\begin{split} z_1^\top Y_{w'} z_1 &\geq (1-\overline{\delta})\lambda_1 \geq (1-\overline{\delta})(3-O(\delta)) \geq 3 - O(\delta + \overline{\delta}) \qquad \text{and} \\ z_1^\top Y_{w'} z_1 &\leq \lambda_1 \alpha_1^2 + \lambda_2 (1-\alpha_1^2) \leq 1 + 2\beta_1^2 + O(\delta). \end{split}$$

Again, this implies that  $\beta_1^2 \ge 1 - O(\delta + \overline{\delta})$ . We can show that  $\min\{\|v_1 - z_1\|_2^2, \|v_1 + z_1\|_2^2\} = O(\overline{\delta} + \delta)$ . By the triangle inequality, we can conclude that  $r^2(z_1) = O(\overline{\delta} + \delta) \le 1/64$ , where the last inequality is obtained by choosing sufficiently small  $\delta$  and  $\overline{\delta}$ . Therefore, there exists an universal constant  $\epsilon' \ge 0$  such that for all  $0 \le \epsilon \le \epsilon'$ , Algorithm 1 takes  $n = \widetilde{\Omega}(d)$  samples and outputs  $z_1$  such that  $r(z_1) \le 1/8$  with probability at least 0.95.

**Lemma 4.4.** Algorithm 1 runs in time  $\tilde{O}(nd)$ .

**Proof of Lemma 4.4.** Since we have the factorization of the rank-two matrices  $A_i$  and rank-one matrices  $B_i$  for all i = 1, 2, ..., n, and the time to perform a matrix-vector product with  $C_i$  and  $D_i$  is O(d). Therefore, by Lemma 2.1, with  $t_C$  and  $t_D$  to be  $\tilde{O}(nd)$ , Line 3 runs in  $\tilde{O}(nd)$  time. In Line 5, by Lemma 2.2, the top eigenvector of  $Y_{w'}$  can be computed in  $\tilde{O}(n \log d)$  time using power method. Scaling in Line 4 runs in O(n) time. As a result, Algorithm 1 runs in  $\tilde{O}(nd)$  time.

<sup>275</sup> We can directly combine Lemma 4.3 and Lemma 4.4 to finish the proof of Lemma 3.2.

## 276 **5 Robust Gradient Descent**

After the robust spectral initialization in Section 4, we have an initial guess  $z_1 \in \mathbb{R}^d$  that is close to the ground truth x or -x. Without loss of generality, we can assume that  $z_1$  is closer to x than to -x. In this section, we prove Lemma 3.3: Given an initial guess  $z_1$  with  $||z_1 - x||_2 \le 1/8$ , we can use a robust gradient descent algorithm (Algorithm 2) to recover x to any desire precision  $\Delta > 0$ . It is well-known that gradient descent can achieve geometric convergence rates in non-adversarial settings. We show that Algorithm 2 achieves a similar convergence rate even when the input is  $\epsilon$ -corrupted.

Consider the natural nonconvex formulation:  $\min_{z \in \mathbb{R}^d} \sum_{i=1}^n f_i(z)$  where  $f_i(z) \doteq (\langle a_i, z \rangle^2 - y_i)^2$ . Let  $g_i$  denote the gradient of  $f_i$  with respect to z. Let  $\mathcal{D}_z$  denote the distribution of  $g_i(z) \in \mathbb{R}^d$  when there is no adversarial corruption. Formally,  $g(z) \sim \mathcal{D}_z$  is distributed as

$$g(z) = \frac{\partial}{\partial z} \left[ \left( \langle a, z \rangle^2 - \langle a, x \rangle^2 \right)^2 \right] = -4 \left( \langle a, z \rangle^2 - \langle a, x \rangle^2 \right) \langle a, z \rangle a \text{ where } a \sim \mathcal{N}(0, I) \text{ .}$$
(7)

- To run gradient descent, we want to estimate the *expected true gradient*  $\mu_z \doteq \mathbb{E} g(z)$  using samples. 286
- The challenge is that the input samples  $\{(a_i, y_i)\}_{i \in [n]}$  are  $\epsilon$ -corrupted, and consequently the gradients 287

 $\{g_i(z)\}_{i\in[n]}$  is an  $\epsilon$ -corrupted set of vectors drawn from  $\mathcal{D}_z$ . To address this, we use robust mean 288 289 estimation algorithms (e.g., [16]) to approximate  $\mu_z$ , the true mean of  $\mathcal{D}_z$ .

Algorithm 2 Robust Gradient Descent

**Input:**  $\epsilon > 0$ , an  $\epsilon$ -corrupted set of n samples  $\{(a_i, y_i)\}_{i \in [n]}$ , initial guess  $z_1 \in \mathbb{R}^d$  with  $||z_1 - x|| \le 1/8$ , and desired precision  $\Delta > 0$ .

**Output:**  $z \in \mathbb{R}^d$  such that  $||z - x||_2 \leq \Delta$  where x is the ground truth.

1: procedure ROBUSTGD( $\epsilon$ ,  $\{(a_i, y_i)\}_{i \in [n]}, z_1, \Delta$ )

2:  $T \leftarrow O(\log(1/\Delta)), \eta \leftarrow 1/300$ 

 $\{N_1, N_2, \cdots, N_T\} \leftarrow$  a random disjoint partition of [n] such that  $|N_t| = n/T$  for all  $t \in [T]$  for  $t = 1, 2, \dots, T$  do 3:

4:

 $\hat{\mu}_{z_t} \leftarrow \text{Robust mean estimation on input } \{g_i(z_t)\}_{i \in N_t} \text{ using Lemma 5.2}$ 5:

6:  $z_{t+1} \leftarrow z_t - \eta \,\widehat{\mu}_{z_t}$ 

7: end for

8: return  $z_{T+1}$ 

9: end procedure

The error guarantee of robust mean estimation algorithms depends on the covariance matrix  $\Sigma_z$  of 290 the distribution  $\mathcal{D}_z$ . The next lemma upper bounds the spectral norm of  $\Sigma_z$ . 291

**Lemma 5.1.** Let  $\mathcal{D}_z$  be the distribution of gradients at z as defined in Equation (7). For any  $z \in \mathbb{R}^d$  with  $||z - x||_2 \leq 1$ , the covariance matrix  $\Sigma_z$  of  $\mathcal{D}_z$  satisfies  $\Sigma_z \preceq O(||z - x||_2^2)I$ . 292 293

The proof of Lemma 5.1 is deferred to Appendix B. Given Lemma 5.1, we can show that robust 294 mean estimation algorithms can approximate  $\mu_z$  with small error. For technical reasons, we randomly 295 partition the input samples  $(a_i, y_i)$  into T subsets, and use one subset in each iteration. With high 296 probability, each partition has at most  $(2\epsilon)$ -fraction of corrupted samples 297

**Lemma 5.2.** Consider any  $z \in \mathbb{R}^d$  with  $||z - x||_2 \leq 1$ . Let  $\mu_z$  be the mean of  $\mathcal{D}_z$  as defined in Equation (7). Fix universal constants c > 0 and  $\epsilon_0 = \Theta(c^2)$ . Let  $2\epsilon < \epsilon_0$  and  $\delta > 0$ . Given  $a(2\epsilon)$ -corrupted set of  $m = \Omega(d \log d/\delta)$  samples drawn from  $\mathcal{D}_z$ , we can compute  $\hat{\mu}_z$  in time 298 299 300  $\widetilde{O}(md\log(1/\delta))$  such that  $\|\widehat{\mu}_z - \mu_z\|_2 \le c \|z - x\|_2$  with probability at least  $1 - O(\delta)$ . 301

*Proof of Lemma 5.2.* Since  $2\epsilon < \epsilon_0$ , we can view the  $(2\epsilon)$ -corrupted set of m samples as  $\epsilon_0$ -corrupted. 302 We need to replace  $2\epsilon$  with  $\epsilon_0$  to reduce the failure probability of Lemma 2.3. This weakens the error 303 guarantee of Lemma 2.3, but the resulting  $\hat{\mu}_z$  is still accurate enough for our algorithm. 304

We apply Lemma 2.3 to the  $\epsilon_0$ -corrupted set of m vectors drawn from  $\mathcal{D}_z$ . By Lemma 5.1, 305 the covariance matrix of  $\mathcal{D}_z$  satisfies  $\Sigma_z \leq O(\|z - x\|_2^2)I$ . Consequently, for sufficiently large  $m = \Theta(d \log d/\delta)$  and sufficiently small  $\epsilon_0 = O(c^2)$ , the error guarantee of Lemma 2.3 is  $O\left(\sqrt{\epsilon_0} + \sqrt{\frac{d}{m\delta}} + \sqrt{\frac{d(\log d + \log(1/\delta))}{m}}\right) \|z - x\|_2 \leq c \|z - x\|_2$ . The success probability is at least 306 307 308  $1 - \delta - \exp(-\epsilon_0 m) = 1 - O(\epsilon_0 m)$ 309

Lemma 5.2 shows that even with a  $(2\epsilon)$ -corrupted set of gradients, the true gradient  $\mu_z$  can be 310 estimated up to an additive error proportional to the distance between z and x. The next lemma shows 311 that such an approximate gradient is sufficient for gradient descent to converge, reducing the distance 312 to the ground truth x by a constant factor in each iteration. 313

**Lemma 5.3.** Suppose in iteration t of Algorithm 2, the current solution  $z_t$  satisfies  $||z_t - x||_2 \le 1/8$ , and the estimated gradient  $\hat{\mu}_{z_t} \in \mathbb{R}^d$  satisfies  $||\hat{\mu}_{z_t} - \mu_{z_t}||_2 \le c ||z_t - x||_2$  for c = 4. Then, we have 314 315  $\|z_{t+1} - x\|_2^2 \le 0.99 \|z_t - x\|_2^2$ 316

Proof Sketch of Lemma 5.3. We provide a proof sketch and defer the full proof to Appendix B. Our 317 objective function is nonconvex (even with infinitely many samples and no corruption). However, 318 when the starting point  $z_1$  is close to a global optimum, it is well-known that gradient descent is 319

well-behaved. More specifically, for any z close to the ground truth x, we can show that the (expected) true gradient  $\mu_z$  aligns with the direction toward x:

$$\langle \mu_z, z - x \rangle \ge 7.5 \, \|z - x\|_2^2$$
 and  $\|\mu_z\|_2 \le 29 \, \|z - x\|_2$ ,

which is sufficient for proving geometric convergence. We can immediately see that this argument is robust to additive error in  $\mu_z$  that is proportional to  $||z - x||_2$ . When  $||\hat{\mu}_z - \mu_z||_2 \le c ||z - x||_2$ ,

$$\langle \hat{\mu}_z, z - x \rangle \ge (7.5 - c) \, \|z_t - x\|_2^2 \text{ and } \|\hat{\mu}_{z_t}\|_2 \le (29 + c) \, \|z_t - x\|_2$$

When c < 7.5, we can choose an appropriate step size  $\eta$  such that the distance between  $z_t$  and xdecreases by a constant factor in each iteration.

We are now ready to prove Lemma 3.3, which states the performance guarantee, sample complexity, runtime, and success probability of Algorithm 2. We restate Lemma 3.3 before proving it.

Lemma 3.3 (Robust Gradient Descent). Consider the setting of Problem 1.3. Let  $\Delta > 0$  be the desired precision. Let  $0 < \epsilon < \epsilon_0$  for some sufficiently small universal constant  $\epsilon_0$ . Given an  $\epsilon$ -corrupted set of  $n = \tilde{\Omega}(d\log^2(1/\Delta))$  samples and an initial guess  $z_1$  such that  $r(z_1) = \min(||z_1 - x||_2, ||z_1 + x||_2) \le 1/8$ , we can compute a vector  $z \in \mathbb{R}^d$  in time  $\tilde{O}(nd)$  such that  $r(z) < \Delta$  with probability at least 0.95.

Proof of Lemma 3.3. First, we prove the success probability of Algorithm 2. Algorithm 2 can fail in two ways: (i) if some  $N_t$  has more than  $(2\epsilon)$ -fraction of corrupted samples, or (ii) if Lemma 5.2 fails in some iteration t. The probability of event (i) is at most 0.01 for our choice of n, which follows from a standard application of Hoeffding's inequality and a union bound over T iterations. For event (ii), we choose a sufficiently small  $\delta = O(1/T)$  in Lemma 5.2, so each robust gradient estimation fails with probability at most  $O(\delta) = 0.01/T$ , and overall the probability of event (ii) is at most 0.01. For the rest of the proof, we assume these bad events do not happen.

Next, we show the correctness of Algorithm 2. Without loss of generality, we can assume that  $z_1$  is closer to the ground truth x than to -x, which implies  $||z_1 - x||_2 \le 1/8$ . By Lemma 5.2, we can obtain an approximation  $\hat{\mu}_{z_1}$  of the true gradient  $\mu_{z_1}$  at  $z_1$  such that  $||\hat{\mu}_{z_1} - \mu_{z_1}||_2 \le c ||z_1 - x||_2$ . Then by Lemma 5.3, we know that  $||z_2 - x||_2 \le 0.99 ||z_1 - x||_2$  after one step of gradient descent. We can iteratively apply these two lemmas to show that, after  $T = O(\log(1/\Delta))$  iterations, we have  $||z_{T+1} - x||_2 \le \Delta$ . One technical issue is that in iteration t, we need to use a fresh subset of samples  $N_t$ . By the principle of deferred decisions, we can view  $(a_i, y_i)_{i \in N_t}$  as being generated (and corrupted) after  $z_t$  is chosen, which forms a  $(2\epsilon)$ -corrupted set of gradients at  $z_t$ .

Finally, we analyze the sample complexity and runtime of Algorithm 2. Algorithm 2 requires in total  $n = mT = \Omega(d \log d \log^2(1/\Delta))$  samples. A random partition can be computed in O(n) time by shuffling the input. In each iteration, the *m* gradients in  $N_t$  can be computed using Equation (7) in time O(md), and  $z_t$  can be updated in time O(d). By Lemma 5.2, the true gradient can be robustly estimated in time  $\widetilde{O}(md \log T) = \widetilde{O}(md \log \log(1/\Delta))$ . The overall runtime of the algorithm is  $\widetilde{O}(n + (md \log \log(1/\Delta))T) = \widetilde{O}(nd \log \log(1/\Delta)) = \widetilde{O}(nd)$ .

## **354 6 Conclusions and Future Directions**

In this paper, our main conceptual contribution is to propose and study the outlier-robust phase retrieval problem, where a constant fraction of the input data is corrupted. Notably, we allow corruption in both the sampled frequencies  $a_i \in \mathbb{R}^d$  and the intensity measurements  $y_i \in \mathbb{R}$ . Our main technical contribution is the design and analysis of a near-sample-optimal and nearly-linear-time algorithm that solves this problem with provably guarantees.

An immediate technical question is whether our sample complexity can be tightened by removing 360 some  $\log(1/\Delta)$  factors. One potential approach is to open robust mean estimation algorithms instead 361 of using them in a black-box manner. One could examine the stability conditions that these algorithms 362 require, and see if these stability conditions can be proved without partitioning the samples and using 363 fresh samples in each iteration. More broadly, we believe our framework can be used to develop 364 outlier-robust algorithms for other tractable nonconvex problems, by first finding an initial solution in 365 a saddle-free region near a global optimum and then converging to this global optimum using robust 366 gradient descent. 367

## 368 References

- [1] A. S. Bandeira, N. Boumal, and V. Voroninski. On the low-rank approach for semidefinite
   programs arising in synchronization and community detection. In *Conference on learning theory*, pages 361–382. PMLR, 2016.
- [2] S. Bhojanapalli, B. Neyshabur, and N. Srebro. Global optimality of local search for low rank matrix recovery. *Advances in Neural Information Processing Systems*, 29, 2016.
- [3] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory*, 61(4):1985–2007, 2015a.
- [4] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns.
   *Applied and Computational Harmonic Analysis*, 39(2):277–299, Sept. 2015b. ISSN 1063-5203.
- [5] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion.
   *SIAM Rev.*, 57(2):225–251, 2015c.
- [6] J. Chen, L. Wang, X. Zhang, and Q. Gu. Robust Wirtinger Flow for Phase Retrieval with Arbitrary Corruption, Jan. 2018.
- [7] Y. Chen and E. Candes. Solving Random Quadratic Systems of Equations Is Nearly as Easy as
   Solving Linear Systems. In *Advances in Neural Information Processing Systems*, volume 28.
   Curran Associates, Inc., 2015.
- [8] Y. Cheng and R. Ge. Non-convex matrix completion against a semi-random adversary. In
   *Conference On Learning Theory*, pages 1362–1394. PMLR, 2018.
- [9] Y. Cheng, I. Diakonikolas, and R. Ge. High-dimensional robust mean estimation in nearly-linear
   time. In *Proceedings of the 30th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages
   2755–2771. SIAM, 2019.
- [10] Y. Cherapanamjeri, S. Mohanty, and M. Yau. List decodable mean estimation in nearly linear
   time. In S. Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 141–148. IEEE, 2020.
- [11] C. Dainty and J. R. Fienup. Phase retrieval and image reconstruction for astronomy. *Image recovery: theory and application*, 231:275, 1987.
- I. Diakonikolas and D. M. Kane. *Algorithmic High-Dimensional Robust Statistics*. Cambridge
   University Press, 2023.
- I. Diakonikolas, G. Kamath, D. M. Kane, J. Li, A. Moitra, and A. Stewart. Robust estimators in high dimensions without the computational intractability. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*, pages 655–664. IEEE Computer Soc., Los Alamitos, CA, 2016.
- [14] I. Diakonikolas, G. Kamath, D. Kane, J. Li, J. Steinhardt, and A. Stewart. Sever: A robust
   meta-algorithm for stochastic optimization. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1596–1606. PMLR, 09–15 Jun 2019.
- [15] Y. Dong, S. Hopkins, and J. Li. Quantum Entropy Scoring for Fast Robust Mean Estimation and
   Improved Outlier Detection. In *Advances in Neural Information Processing Systems*, volume 32.
   Curran Associates, Inc., 2019.
- [16] Y. Dong, S. B. Hopkins, and J. Li. Quantum entropy scoring for fast robust mean estimation and
   improved outlier detection. In *Proc. 33rd Advances in Neural Information Processing Systems* (*NeurIPS*), 2019.
- [17] J. R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3(1):27–29, July 1978. ISSN 1539-4794.
- [18] J. R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769,
   Aug. 1982. ISSN 2155-3165.

- [19] X. Gao and Y. Cheng. Robust matrix sensing in the semi-random model. *Proceedings of the* 37th Conference on Neural Information Processing Systems (NeurIPS), 2023.
- [20] R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping from saddle points—online stochastic gradient
   for tensor decomposition. In *Conference on Learning Theory*, pages 797–842, 2015.
- [21] R. Ge, J. D. Lee, and T. Ma. Matrix completion has no spurious local minimum. In Advances in Neural Information Processing Systems, pages 2973–2981, 2016.
- 421 [22] R. W. Gerchberg. A practical algorithm for the determination of phase from image and 422 diffraction plane pictures. *Optik*, Jan. 1972.
- <sup>423</sup> [23] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics. The* <sup>424</sup> *approach based on influence functions.* Wiley New York, 1986.
- P. Hand and V. Voroninski. Corruption robust phase retrieval via linear programming. *CoRR*, abs/1612.03547, 2016.
- 427 [25] P. J. Huber and E. M. Ronchetti. *Robust statistics*. Wiley New York, 2009.
- [26] K. Jaganathan, Y. C. Eldar, and B. Hassibi. Phase retrieval: An overview of recent developments.
   *Optical Compressive Imaging*, pages 279–312, 2016.
- 430 [27] R. Kolte and A. Özgür. Phase Retrieval via Incremental Truncated Wirtinger Flow, June 2016.
- [28] K. A. Lai, A. B. Rao, and S. Vempala. Agnostic estimation of mean and covariance. In *focs2016*,
   pages 665–674, 2016.
- 433 [29] G. Lecué, M. Lerasle, and T. Mathieu. Robust classification via MOM minimization. *Mach.* 434 *Learn.*, 109(8):1635–1665, 2020.
- [30] S. Li, Y. Cheng, I. Diakonikolas, J. Diakonikolas, R. Ge, and S. Wright. Robust second-order
   nonconvex optimization and its application to low rank matrix sensing. In *Proc. 37th Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [31] J.-W. Liu, Z.-J. Cao, J. Liu, X.-L. Luo, W.-M. Li, N. Ito, and L.-C. Guo. Phase Retrieval via
   Wirtinger Flow Algorithm and Its Variants. In *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, pages 1–9, July 2019.
- [32] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson. High resolution
   3D X-ray diffraction microscopy. *Physical review letters*, 89(8):088303, 2002.
- [33] R. P. Millane. Phase retrieval in crystallography and optics. JOSA A, 7(3):394–411, 1990.
- [34] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Proc.* 27th Advances in Neural Information Processing Systems (NeurIPS), pages 2796–2804, 2013.
- [35] A. Prasad, A. S. Suggala, S. Balakrishnan, and P. Ravikumar. Robust estimation via robust
   gradient estimation. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*,
   82(3):601–627, 2020.
- <sup>449</sup> [36] W. H. Robert. Phase problem in crystallography. *JOSA a*, 10(5):1046–1055, 1993.
- [37] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval
   with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*,
   32(3):87–109, 2015.
- [38] M. Soltanolkotabi. Structured signal recovery from quadratic measurements: Breaking sample
   complexity barriers via nonconvex optimization. *IEEE Trans. Inf. Theory*, 65(4):2374–2400,
   2019.
- [39] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere i: Overview and the
   geometric picture. *IEEE Trans. Inf. Theor.*, 63(2):853–884, 2 2017.
- [40] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. *Found. Comput. Math.*,
   18(5):1131–1198, 2018.

- [41] L. Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. University of California, Berkeley, https://people. eecs. berkeley. edu/luca/books/expanders-2016. pdf, 2017.
- [42] J. Tukey. Mathematics and picturing of data. In *Proceedings of ICM*, volume 6, pages 523–531, 1975.
- [43] V. Voroninski. *PhaseLift: A Novel Methodology for Phase Retrieval*. PhD thesis, UC Berkeley, 2013.
- [44] J. Wright and Y. Ma. *High-dimensional data analysis with low-dimensional models: Principles, computation, and applications.* Cambridge University Press, 2022.
- <sup>468</sup> [45] H. Zhang, Y. Chi, and Y. Liang. Provable non-convex phase retrieval with outliers: Median <sup>469</sup> truncated wirtinger flow. In *International conference on machine learning*, pages 1022–1031.
- 470 PMLR, 2016.

## 471 A Omitted Proofs in Section 4

**Lemma A.1.** [Lemma 4.2, Formal]. For any  $\delta_0 > 0$ , there exists constants  $\epsilon_0, c > 0$  such that when  $n > c \cdot d \log d$  and we are given a set of  $n \epsilon$ -corrupted samples, where  $0 \le \epsilon \le \epsilon_0$ , then with probability at least 0.98, it holds

$$\forall w \in \Delta_{n,2\epsilon}, \left\| Y_{G,w} - (I + 2xx^{\top}) \right\|_2 \le \delta_0.$$
(8)

Proof of Lemma A.1. We recall the definition of  $Y_{G,w} = \sum_{i \in G} w_i y_i a_i a_i^{\top}$ . Let  $\ell = \epsilon \cdot n$  and let  $\{(a_{n+i}, y_{n+i})\}_{i=1}^{\ell}$  be the set of samples that were removed by the  $\epsilon$ -corruption adversary. Let  $G' = G \cup \{n+1, \dots, n+\ell\}, n' = n+\ell$ , and  $\epsilon' = \epsilon/(1+\epsilon)$ . Note that without loss of generality, we can assume that  $|G| = (1-\epsilon)n$  and  $|G'| = (1-\epsilon')n' = n$ .

479 We define a mapping  $\sigma: \Delta_{n,2\epsilon} \to \Delta_{n',3\epsilon'}$  such that

$$\sigma(w)_i = \begin{cases} w_i & i \in [n] \\ 0 & \text{otherwise} \end{cases}$$
(9)

In other words, all the weights are the same for the samples with index in the set [n], and are equal to 0 for the samples removed by the adversary. We can verify that  $\sigma(w) \in \Delta_{n',3\epsilon'}$  for all  $w \in \Delta_{n,2\epsilon}$ since  $\sigma(w)_i \leq w_i \leq 1/(1-2\epsilon)n = 1/(1-3\epsilon')n'$  for all  $i \in [n']$ , and  $\|\sigma(w)\|_1 = \|w\|_1 = 1$ . Furthermore, we have  $Y_{G,w} = Y_{G',\sigma(w)}$  for all  $w \in \Delta_{n,2\epsilon}$ . We denote with  $w^* \in \Delta_{n',3\epsilon'}$  the desired uniform weighting of the samples with index in G', i.e.,  $w_i^* = \frac{1}{(1-\epsilon')n'} \mathbb{1}_{i \in G'}$ .

485 It suffices to show both  $\|Y_{G',w^*} - (I + 2xx^\top)\|_2 \le \delta_0/2$  and  $\|Y_{G',\sigma(w)-w^*}\|_2 \le \delta_0/2$ .

486 By triangle inequality, for any  $w \in \Delta_{n,2\epsilon}$ , it holds

$$\left\|Y_{G',\sigma(w)} - (I + 2xx^{\top})\right\|_{2} \le \left\|Y_{G',w^{*}} - (I + 2xx^{\top})\right\|_{2} + \left\|Y_{G',\sigma(w)-w^{*}}\right\|_{2},$$

487 Thus, it suffices to show both  $||Y_{G',w^*} - (I + 2xx^{\top})||_2 \le \delta_0/2$  and  $||Y_{G',\sigma(w)-w^*}||_2 \le \delta_0/2$ .

488 We upper bound the first term. By using the definition of  $w^*$ , note that

$$\begin{split} \left\| Y_{G',\sigma(w)} - (I + 2xx^{\top}) \right\|_{2} &= \left\| \sum_{i \in G'} w_{i}^{*} y_{i} a_{i} a_{i}^{\top} - (I + 2xx^{\top}) \right\|_{2} \\ &= \left\| \sum_{i \in G'} \frac{1}{|G'|} y_{i} a_{i} a_{i}^{\top} - (I + 2xx^{\top}) \right\|_{2} \end{split}$$

Since  $\mathbb{E}(y_i a_i a_i^{\top}) = I + 2xx^{\top}$  for any  $i \in G'$ , we can use a concentration inequality to upper bound this term. By Lemma A.2, as long as  $n \ge c_1(\delta_0/2) \cdot d \log d$ , with probability at least 0.995, we have

$$\left\|Y_{G',w^*} - (I + 2xx^{\top})\right\|_2 \le \delta_0/2 \quad . \tag{10}$$

It remains to show a high-probability upper bound to the second term  $||Y_{G',w^*-\sigma(w)}||_2 \leq \delta_0/2$ . The first observation is that the weighting  $w^*$  and  $\sigma(w)$  for any  $w \in \Delta_{n,2\epsilon}$  cannot too different. In particular, we can show the following upper bound:

$$\sum_{i \in G'} |w_i^* - \sigma(w)_i| \le \sum_{i=1}^{n'} |w_i^* - \sigma(w)_i| \le \sup_{w, w' \in \Delta_{n', 3\epsilon'}} \sum_{i=1}^{n'} |w_i - w_i'|$$
(11)

We observe that  $\Delta_{n',3\epsilon'}$  can be seen as the convex combination of all possible uniform weighting over subsets of  $n'(1-3\epsilon')$  samples. Thus, the maximum distance will be between two points of the convex hull, and we can upper bound (11) as:

$$\sum_{i \in G'} |w_i^* - \sigma(w)_i| \le \sup_{w, w' \in \Delta_{n', 3\epsilon'}} \sum_{i=1}^{n'} |w_i - w_i'| \le \frac{6\epsilon' n}{n'(1 - 3\epsilon')} \le 6\epsilon \quad .$$
(12)

497 For a fixed unit vector  $z \in \mathbb{S}^{d-1}$  with z = px + qu where  $u \in \mathbb{S}^{d-1}$  and  $\langle u, x \rangle = 0$ , we have:

$$\begin{split} \max_{w \in \Delta_{n,2\epsilon}} \left| z^{\top} Y_{G',w^* - \sigma(w)} z \right| &= \max_{w} \left| \sum_{i \in G'} (w_i^* - \sigma(w)_i) \left\langle a_i, x \right\rangle^2 \left\langle a_i, z \right\rangle^2 \right| \\ &= \max_{w} \left| \sum_{i \in G'} (w_i^* - \sigma(w)_i) \left\langle a_i, x \right\rangle^2 (p \left\langle a_i, x \right\rangle + q \left\langle a_i, u \right\rangle)^2 \right| \\ &\leq 2 \max_{w} \left| \sum_{i \in G'} (w_i^* - \sigma(w)_i) (\left\langle a_i, x \right\rangle^4 + \left\langle a_i, x \right\rangle^2 \left\langle a_i, u \right\rangle^2) \right| \\ &\leq 2 \max_{w} \sum_{i \in G'} |w_i^* - \sigma(w)_i| \left\langle a_i, x \right\rangle^4 \\ &+ 2 \max_{w} \sum_{i \in G'} |w_i^* - \sigma(w)_i| \left\langle a_i, x \right\rangle^2 \left\langle a_i, u \right\rangle^2 \ . \end{split}$$

For ease of notation, let  $\beta_i \doteq |w_i^* - \sigma(w)_i|$ . Observe that  $0 \le \beta_i \le \frac{1}{(1-2\epsilon)n}$  for all i, and  $\sum_i \beta_i \le 6\epsilon$ due to (11). We have:

$$\max_{w \in \Delta_{n,2\epsilon}} \left| z^{\top} Y_{G',w^* - \sigma(w)} z \right| \leq 2 \max_{\beta: \sum_{i \in G'} \beta_i \leq 6\epsilon \text{ and } 0 \leq \beta_i \leq \frac{1}{(1 - 2\epsilon)n}} \sum_{i \in G'} \beta_i \langle a_i, x \rangle^4 
+ 2 \max_{\beta: \sum_{i \in G'} \beta_i \leq 6\epsilon \text{ and } 0 \leq \beta_i \leq \frac{1}{(1 - 2\epsilon)n}} \sum_{i \in G'} \beta_i \langle a_i, x \rangle^2 \langle a_i, u \rangle^2 
\leq \frac{2}{(1 - 2\epsilon)n} \max_{L \subseteq G', |L| = 6\epsilon n} \sum_{i \in L} \langle a_i, x \rangle^4 
+ \frac{2}{(1 - 2\epsilon)n} \max_{L \subseteq G', |L| = 6\epsilon n} \sum_{i \in L} \langle a_i, x \rangle^2 \langle a_i, u \rangle^2.$$
(13)

Inequality (13) follows by assigning the maximum possible  $\beta_i$  to the largest entries of the sum until we hit the budget  $6\epsilon$  due to (11).

We bound  $\max_L \sum_{i \in L} \langle a_i, x \rangle^4$  first. Let  $X_i = \langle a_i, x \rangle \sim \mathcal{N}(0, 1)$  for  $i \in G'$ , and define the threshold function  $h_r(z) = \begin{cases} 0, & z \leq r \\ z, & z > r \end{cases}$  for  $r = C^2 \cdot \ln^2(1/\epsilon)$  with constant C > 0 to be determined. Note that  $z \leq r + h_r(z)$  for all z > 0. Therefore,

$$\begin{split} \max_{L \subseteq G', |L| = 6\epsilon n} \frac{1}{n} \sum_{i \in L} X_i^4 &\leq \max_L \frac{1}{n} \sum_{i \in L} r + \max_L \frac{1}{n} \sum_{i \in L} h_r(X_i^4) \\ &\leq 6\epsilon r + \frac{1}{n} \sum_{i \in G'} h_r(X_i^4). \end{split}$$

Then, we consider to bound  $\exp\left(\sum_{i\in G'} c \cdot h_r(X_i^4)\right)$  for some c > 0 to be determined. For any  $i \in G'$  and  $z \ge 1$ , with C = 6, for all  $\epsilon > 0$ , we have

$$\mathbf{Pr}\left[\exp\left(c \cdot h_r(X_i^4)\right) \ge z\right] \le \mathbf{Pr}\left[h_r(X_i^4) \ge 0\right] = \mathbf{Pr}\left[X_i \ge r^{1/4}\right] \le \exp(-\sqrt{r}/2)$$
$$\le \exp(\ln \epsilon \cdot C/2)$$
$$\le \epsilon^3.$$

507 At the same time, with c < 1/200, for all  $z \ge 1$ , we have

$$\mathbf{Pr}\left[\exp\left(c \cdot h_r(X_i^4)\right) \ge z\right] \le \mathbf{Pr}\left[\exp(c \cdot X_i^4) \ge z\right] \le \mathbf{Pr}\left[X_i^4 \ge \frac{\ln z}{c}\right]$$
$$\le \exp\left(-\sqrt{\frac{\ln z}{c}}/2\right)$$
$$\le z^{-3}.$$

Therefore,  $\Pr\left[\exp\left(c \cdot h_r(X_i^4)\right) \ge z\right] \le \min\{z^{-3}, \epsilon^3\}$ , and for all  $\epsilon < 1/2$ , we have

$$\mathbf{E}\left[\exp\left(c \cdot h_r(X_i^4)\right)\right] = \int_0^\infty \mathbf{Pr}\left[\exp\left(c \cdot h_r(X_i^4)\right) \ge z\right] dz$$
$$\leq \int_0^1 1 dz + \int_1^{1/\epsilon} \epsilon^3 dz + \int_{1/\epsilon}^\infty z^{-3} dz$$
$$= 1 + \epsilon^2 - \epsilon^3 + \frac{1}{2}\epsilon^2$$
$$\leq 1 + \epsilon^2$$
$$< \exp(\epsilon^2).$$

Since  $\{X_i\}_{i \in G'}$  are independent, we have

$$\mathbf{E}\left[\exp\left(\sum_{i\in G'}c\cdot h_r(X_i^4)\right)\right] = \mathbf{E}\left[\prod_{i\in G'}\exp\left(c\cdot h_r(X_i^4)\right)\right] \le \exp(\epsilon^2 n).$$

Finally, by Markov's inequality, for any constant  $\delta_1 > 0$ , as long as  $\epsilon \leq \sqrt{\delta_1 c}$ , we have

$$\mathbf{Pr}\left[\sum_{i\in G'} h_r(X_i^4) \ge 2\delta_1 n\right] = \mathbf{Pr}\left[\exp\left(\sum_{i\in G'} c \cdot h_r(X_i^4)\right) \ge \exp(2\delta_1 cn)\right]$$
$$\le \exp(\epsilon^2 n - 2\delta_1 cn) \le \exp(-\delta_1 cn).$$

As a result, with sufficiently large  $n \ge c_2(\delta_1) \cdot d \log d$  and sufficiently small  $\epsilon$  such that  $6\epsilon r \le \delta_1$ ,

$$\mathbf{Pr}\left[\epsilon r + \frac{1}{n}\sum_{i\in G'}h_r(X_i^4) \ge 3\delta_1\right] \le \exp(-\delta_1 cn) \le 0.995 \cdot 9^{-d}.$$

Also,  $\max_L \sum_{i \in L} \langle a_i, x \rangle^2 \langle a_i, u \rangle^2$  has a similar tail bound and therefore can be bounded in the same way. We can then bound the operator norm via an epsilon-net argument. Set  $\delta_1 = \delta_0/24$ . By an 1/4-net on  $S^{d-1}$  with  $|\mathcal{N}| \leq 9^d$ , we have that

$$\mathbf{Pr}\left[\max_{z\in\mathbb{S}^{d-1}}\left|z^{\top}Y_{G',w^*-\sigma(w)}z\right| \ge \delta_0/2\right] \le 9^d \cdot 0.99 \cdot 9^{-d} \le 0.99.$$

By combining the above inequality with Equation (10), we can conclude that, for any  $\delta_0 > 0$ , there exists  $\epsilon_0 > 0$ , such that when  $n \ge \max\{c_1(\delta_0), c_2(\delta_0)\} \cdot d \log d$  and  $0 \le \epsilon \le \epsilon_0$ , with probability at least 0.98,

$$\forall w \in \Delta_{n,2\epsilon}, \left\| Y_{G,w} - (I + 2xx^{\top}) \right\|_2 = \left\| Y_{G',\sigma(w)} - (I + 2xx^{\top}) \right\|_2 \le \delta_0.$$

Therefore, with probability at least 0.98, we have  $\|Y_{G,w} - (I + 2xx^{\top})\|_2 \le \delta_0$  for all  $w \in \Delta_{n,2\epsilon}$ .

## 520 A.1 Concentration Inequalities

521 For the undisturbed samples, we have the following concentration result.

Lemma A.2 ([3] Section A.4.2). Let  $x \in \mathbb{R}^d$ . For any  $\delta > 0$ , there exists a constant  $C(\delta) > 0$ such that when  $n > C \cdot d \log d$  and we are given a set of n samples  $\{(a_i, y_i)\}_{i=1}^n$  with  $a_i \sim \mathcal{N}(0, I)$ 

independently and  $y_i = \langle a_i, x \rangle^2$  for all  $i \in [n]$ , then with probability at least 0.99, it holds

$$\| \quad n = \|$$

$$\left\|\frac{1}{n}\sum_{i=1}^{n}y_{i}a_{i}a_{i}^{\top}-(I+2xx^{\top})\right\|_{2}\leq\delta.$$

## 525 **B** Omitted Proofs in Section 5

Lemma 5.1. Let  $\mathcal{D}_z$  be the distribution of gradients at z as defined in Equation (7). For any  $z \in \mathbb{R}^d$ with  $||z - x||_2 \leq 1$ , the covariance matrix  $\Sigma_z$  of  $\mathcal{D}_z$  satisfies  $\Sigma_z \preceq O(||z - x||_2^2)I$ .

<sup>528</sup> Proof of Lemma 5.1. Recall that  $g \sim D_z$  is distributed as

$$g = \frac{\partial}{\partial z} \left[ \left( \langle a, z \rangle^2 - \langle a, x \rangle^2 \right)^2 \right]$$
  
=  $-4 \left( \langle a, z \rangle^2 - \langle a, x \rangle^2 \right) \langle a, z \rangle a$  where  $a \sim \mathcal{N}(0, 1)$ .

Because  $\mathbf{E}_{g \sim \mathcal{D}_z}[g] = \mu_z$ , the spectral norm of  $\Sigma_z$  can be upper bounded as follows:

$$\left\|\Sigma_{z}\right\|_{2} = \left\| \mathbf{\underline{E}}_{g \sim \mathcal{D}_{z}} \left[ gg^{\top} \right] - \mu_{z} \mu_{z}^{\top} \right\|_{2} \le \left\| \mathbf{\underline{E}}_{g \sim \mathcal{D}_{z}} \left[ gg^{\top} \right] \right\|_{2}$$

<sup>530</sup> Consequently, it suffices to upper bound  $\|\mathbf{E}_{g\sim \mathcal{D}_z} [gg^{\top}]\|_2$ . Let h = z - x.

$$\begin{split} \left\| \mathbf{E}_{g \sim \mathcal{D}_{z}} \left[ gg^{\top} \right] \right\|_{2} &= \max_{\|v\|_{2}=1} v^{\top} \mathbf{E}_{g \sim \mathcal{D}_{z}} \left[ gg^{\top} \right] v = \max_{\|v\|_{2}=1} \mathbf{E}_{g} \left[ \langle g, v \rangle^{2} \right] \\ &= 16 \max_{\|v\|_{2}=1} \mathbf{E}_{a \sim \mathcal{N}(0,I)} \left[ \langle \langle a, z \rangle^{2} - \langle a, x \rangle^{2} \rangle^{2} \langle a, z \rangle^{2} \langle a, v \rangle^{2} \right] \\ &= 16 \max_{\|v\|_{2}=1} \mathbf{E}_{a \sim \mathcal{N}(0,I)} \left[ \langle a, h \rangle^{2} \langle a, 2x + h \rangle^{2} \langle a, x + h \rangle^{2} \langle a, v \rangle^{2} \right] \\ &\leq 16 \max_{\|v\|_{2}=1} \left( \mathbf{E}_{a} \left[ \langle a, h \rangle^{8} \right] \mathbf{E}_{a} \left[ \langle a, 2x + h \rangle^{8} \right] \mathbf{E}_{a} \left[ \langle a, x + h \rangle^{8} \right] \mathbf{E}_{a} \left[ \langle a, v \rangle^{8} \right] \right)^{1/4} \\ &= 16 \max_{\|v\|_{2}=1} \left( 105^{4} \|h\|_{2}^{8} \|2x + h\|_{2}^{8} \|x + h\|_{2}^{8} \|v\|_{2}^{8} \right)^{1/4} \\ &= (16 \cdot 105) \|h\|_{2}^{2} \|2x + h\|_{2}^{2} \|x + h\|_{2}^{2} = O(\|h\|_{2}^{2}) \ . \end{split}$$

The first inequality is due to Cauchy-Schwarz inequality. The last step uses the fact that  $||x||_2 = 1$ and  $||h||_2 \le 1$ .

Lemma 5.3. Suppose in iteration t of Algorithm 2, the current solution  $z_t$  satisfies  $||z_t - x||_2 \le 1/8$ , and the estimated gradient  $\hat{\mu}_{z_t} \in \mathbb{R}^d$  satisfies  $||\hat{\mu}_{z_t} - \mu_{z_t}||_2 \le c ||z_t - x||_2$  for c = 4. Then, we have  $||z_{t+1} - x||_2^2 \le 0.99 ||z_t - x||_2^2$ .

536 Proof of Lemma 5.3. Recall that  $g \sim D_z$  is distributed as

$$g = \frac{\partial}{\partial z} \left[ \left( \langle a, z \rangle^2 - \langle a, x \rangle^2 \right)^2 \right] \text{ where } a \sim \mathcal{N}(0, I)$$

537 We can compute the mean of  $D_z$  using moments of Gaussian:

$$\mu_{z} = \mathop{\mathbb{E}}_{g \sim \mathcal{D}_{z}} g = \left( 12 \left\| z \right\|_{2}^{2} - 4 \left\| x \right\|_{2}^{2} \right) z - 8 \langle x, z \rangle x .$$
(14)

<sup>538</sup> Consider one step of gradient descent in Algorithm 2:  $z_{t+1} = z_t - \eta \hat{\mu}_{z_t}$ , where  $\hat{\mu}_{z_t}$  is an approximate <sup>539</sup> gradient. We have

$$\|z_{t+1} - x\|_2^2 = \|z_t - \eta \widehat{\mu}_{z_t} - x\|_2^2 = \|z_t - x\|_2^2 - 2\eta \langle \widehat{\mu}_{z_t}, z_t - x \rangle + \eta^2 \langle \widehat{\mu}_{z_t}, \widehat{\mu}_{z_t} \rangle$$

- To prove convergence, we need to lower bound  $\langle \hat{\mu}_{z_t}, z_t x \rangle$  and upper bound  $\langle \hat{\mu}_{z_t}, \hat{\mu}_{z_t} \rangle$ .
- We write  $z = z_t$  and h = z x to simplify notation. We can substitute z = x + h in Equation (14):

$$\mu_z = \left(16\langle x, h \rangle + 12 \|h\|_2^2\right) x + \left(8 \|x\|_2^2 + 24\langle x, h \rangle + 12 \|h\|_2^2\right) h \,.$$

- Recall the assumptions of this lemma:  $||x||_2 = 1$ ,  $||h||_2 \le 1/8$ , and  $||\hat{\mu}_z \mu_z||_2 \le c ||h||_2$ .
- 543 First we lower bound  $\langle \hat{\mu}_z, h \rangle$ .

$$\begin{split} \langle \widehat{\mu}_{z}, h \rangle &= \langle \mu_{z}, h \rangle + \langle \widehat{\mu}_{z} - \mu_{z}, h \rangle \\ &= 16 \langle x, h \rangle^{2} + 36 \langle x, h \rangle \, \|h\|_{2}^{2} + 8 \, \|x\|_{2}^{2} \, \|h\|_{2}^{2} + 12 \, \|h\|_{2}^{4} + \langle \widehat{\mu}_{z} - \mu_{z}, h \rangle \\ &\geq -\frac{81}{4} \, \|h\|_{2}^{4} + 8 \, \|x\|_{2}^{2} \, \|h\|_{2}^{2} + 12 \, \|h\|_{2}^{4} - \langle \widehat{\mu}_{z} - \mu_{z}, h \rangle \\ &\geq (8 - \frac{33}{256} - c) \, \|h\|_{2}^{2} \\ &\geq (7.5 - c) \, \|h\|_{2}^{2} \, . \end{split}$$

The first inequality uses the fact that  $16\langle x,h\rangle^2 + 36\langle x,h\rangle \|h\|_2^2$  is a second-order polynomial of  $\langle x,h\rangle$ , which has minimum value  $-\frac{81}{4} \|h\|_2^4$  for all  $\langle x,h\rangle \in \mathbb{R}$ .

Next we upper bound  $\langle \hat{\mu}_z, \hat{\mu}_z \rangle$  using the triangle inequality.

$$\begin{split} \|\widehat{\mu}_{z}\|_{2} &\leq \|\mu_{z}\|_{2} + \|\widehat{\mu}_{z} - \mu_{z}\|_{2} \\ &\leq \left(16\langle x, h \rangle + 12 \|h\|_{2}^{2}\right) \|x\|_{2} + \left(8 \|x\|_{2}^{2} + 24\langle x, h \rangle + 12 \|h\|_{2}^{2}\right) \|h\|_{2} + c \|h\|_{2} \\ &\leq \left(16 + \frac{12}{8} + 8 + \frac{24}{8} + \frac{12}{64} + c\right) \|h\|_{2} \\ &\leq \left(29 + c\right) \|h\|_{2} \ . \end{split}$$

547 Putting everything together, we have

$$\begin{aligned} \|z_{t+1} - x\|_2^2 &= \|z_t - x\|_2^2 - 2\eta \langle \widehat{\mu}_{z_t}, z_t - x \rangle + \eta^2 \langle \widehat{\mu}_{z_t}, \widehat{\mu}_{z_t} \rangle \\ &\leq \left[ 1 - 2(7.5 - c)\eta + (29 + c)^2 \eta^2 \right] \|z_t - x\|_2^2 . \end{aligned}$$

548 Choosing c = 4 and  $\eta = 1/300$  gives that  $||z_{t+1} - x||_2^2 \le 0.99 ||z_t - x||_2^2$ .

## 549 C Counter-examples

Prior robust phase retrieval algorithms [24, 45] focus on the setting where the observations  $y_i$ 's are subject to adversarial perturbation while the measuring vectors  $a_i$ 's are independently sampled from the Gaussian distribution. The Median Truncated Wirtinger Flow Algorithm [45] first initialize  $z^{(0)}$  by the spectral method, calculating  $z^{(0)}$  as the top eigenvector of  $Y := \frac{1}{m} \sum_{i=1}^{m} y_i a_i a_i^{\top} \mathbb{1}_{|y_i| \le \alpha^2 \mod\{\{y_i\}_{i=1}^m\}}$  using a truncated set of samples, where the threshold is determined by  $med(\{y_i\}_{i=1}^m)$ , the median over all  $y_i$ 's. As long as the fraction of of outliers is not too large and the sample complexity is large enough, the initialization is guaranteed to be within a small neighborhood of the ground truth.

In this section, we present a counter-example where robust phase retrieval algorithms [24, 45] can be insufficient when directly applied to the  $\epsilon$ -corruption phase retrieval problem.

Let  $x \in \mathbb{S}^{d-1}$  be the ground truth unit vector. Here we construct an  $\epsilon$ -corruption adversary that can manipulate the top eigenvector of the empirical covariance matrix  $Y = \sum_{i=1}^{n} y_i a_i a_i^{\top}$ , even when  $y_i$ are accurately calculated as  $y_i = (a_i^{\top} x)^2$ .

Let  $u \in \mathbb{S}^{d-1}$  be a unit vector such that  $x^{\top}u = 0$ . Suppose the adversary changes 1% of the  $a_i$ 's to  $a_i = \sqrt{d - 1/25} \cdot u + 1/5x$ , and suppose all the  $y_i$ 's are accurate. In particular, the length of the corrupted  $a_i$ 's is comparable to Gaussian vectors, and the corresponding  $y_i = (a_i^{\top}x)^2 = 1/25$ . Consequently, the median-truncated initialization in [45] will not be able to filter out such  $y_i$ . However, the top eigenvector of  $\mathbf{E}[Y] = \mathbf{E}\left[\sum_{i=1}^n y_i a_i a_i^{\top}\right] = O(d)uu^{\top} + O(\sqrt{d})(ux^{\top} + xu^{\top}) + O(1)(I + 2xx^{\top})$  will be manipulated to u, which is far from the ground truth x.

# 568 NeurIPS Paper Checklist

569	1.	Claims
570		Ouestion: Do the main claims made in the abstract and introduction accurately reflect the
571		paper's contributions and scope?
572		Answer: [Yes]
573		Justification: We provide a formal discussion of our theoretical claim in the first section of
574		the paper, and include a formal statement of our resul in Section 3.
575		Guidelines:
576		• The answer NA means that the abstract and introduction do not include the claims
577		made in the paper.
578		• The abstract and/or introduction should clearly state the claims made, including the
579		contributions made in the paper and important assumptions and limitations. A No or
580		NA answer to this question will not be perceived well by the reviewers.
581		• The claims made should match theoretical and experimental results, and reflect how
582		much the results can be expected to generalize to other settings.
583		• It is fine to include aspirational goals as motivation as long as it is clear that these goals
584		are not attained by the paper.
585	2.	Limitations
586		Question: Does the paper discuss the limitations of the work performed by the authors?
587		Answer: [Yes]
588		Justification: We point out to all the assumptions used in our paper. There are a discussion
589		of the limitations in the conclusion. Our algorithm can only handle a corruption $\epsilon < \epsilon_0$ for
590		some universal constant $\epsilon_0$ . Additionally, it is possible that the dependency on $\Delta$ for the
591		sample complexity can be removed.
592		Guidelines:
593		• The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
594		• The systems are encouraged to error a comprete "Limitations" section in their momen
595		• The authors are encouraged to create a separate Limitations section in their paper.
596		• The paper should point out any strong assumptions and now robust the results are to violations of these assumptions (a g independence assumptions, noiseless settings)
597		model well-specification asymptotic approximations only holding locally). The authors
599		should reflect on how these assumptions might be violated in practice and what the
600		implications would be.
601		• The authors should reflect on the scope of the claims made, e.g., if the approach was
602		only tested on a few datasets or with a few runs. In general, empirical results often
603		depend on implicit assumptions, which should be articulated.
604		• The authors should reflect on the factors that influence the performance of the approach.
605		For example, a facial recognition algorithm may perform poorly when image resolution
606		is low or images are taken in low lighting. Or a speech-to-text system might not be
607		used reliably to provide closed captions for online lectures because it fails to handle
608		technical jargon.
609		• The authors should discuss the computational efficiency of the proposed algorithms
610		and how they scale with dataset size.
611 612		• If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
613		• While the authors might fear that complete honesty about limitations might be used by
614		reviewers as grounds for rejection, a worse outcome might be that reviewers discover
615		limitations that aren't acknowledged in the paper. The authors should use their best
616		judgment and recognize that individual actions in favor of transparency play an impor-
617		tant role in developing norms that preserve the integrity of the community. Reviewers
618		will be specifically instructed to not penalize honesty concerning limitations.
619	3.	Theory Assumptions and Proofs

620 621	Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?
622	Answer: [Yes]
623 624	Justification: We formally describe our problem setting and assumption used. All the proofs are in the paper, either in the main content pages or in the appendix.
625	Guidelines:
626	• The answer NA means that the paper does not include theoretical results.
627	• All the theorems, formulas, and proofs in the paper should be numbered and cross-
628	referenced.
629	• All assumptions should be clearly stated or referenced in the statement of any theorems.
630	• The proofs can either appear in the main paper or the supplemental material, but if
631	they appear in the supplemental material, the authors are encouraged to provide a short
632	proof sketch to provide intuition.
633 634	• Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
635	• Theorems and Lemmas that the proof relies upon should be properly referenced.
636	4. Experimental Result Reproducibility
637	Question: Does the paper fully disclose all the information needed to reproduce the main ex-
638	perimental results of the paper to the extent that it affects the main claims and/or conclusions
639	of the paper (regardless of whether the code and data are provided or not)?
640	Answer: [NA]
641	Justification: The paper is a theory paper, and it has no experiments.
642	Guidelines:
643	• The answer NA means that the paper does not include experiments.
644	• If the paper includes experiments, a No answer to this question will not be perceived
645 646	whether the code and data are provided or not
647	• If the contribution is a dataset and/or model, the authors should describe the steps taken
648	to make their results reproducible or verifiable.
649	• Depending on the contribution, reproducibility can be accomplished in various ways.
650	For example, if the contribution is a novel architecture, describing the architecture fully
651	might suffice, or if the contribution is a specific model and empirical evaluation, it may
652 653	dataset or provide access to the model. In general, releasing code and data is often
654	one good way to accomplish this, but reproducibility can also be provided via detailed
655	instructions for how to replicate the results, access to a hosted model (e.g., in the case
656	of a large language model), releasing of a model checkpoint, or other means that are
657	appropriate to the research performed.
658	• while Neuris does not require releasing code, the conference does require an submis- sions to provide some reasonable avenue for reproducibility, which may depend on the
660	nature of the contribution. For example
661	(a) If the contribution is primarily a new algorithm, the paper should make it clear how
662	to reproduce that algorithm.
663	(b) If the contribution is primarily a new model architecture, the paper should describe
664	the architecture clearly and fully.
665 666	(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce
667	the model (e.g., with an open-source dataset or instructions for how to construct
668	the dataset).
669	(d) We recognize that reproducibility may be tricky in some cases, in which case
670	authors are welcome to describe the particular way they provide for reproducibility.
671 672	In the case of closed-source models, it may be that access to the model is limited in some way (e.g. to registered users) but it should be possible for other researchers
673	to have some path to reproducing or verifying the results.

<ul> <li>Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide instructions on the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> </ul> </li> <li>Guestion: The paper is a theory paper, and it has no experiments.</li> <li>The fult details can be provided either with the code, in appendix,</li></ul>	674	5.	Open access to data and code
<ul> <li>tions to faithfully reproduce the main experimental results, as described in supplemental material?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data, submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized version: (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> </ul> </li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The enswer NA means that the paper does not include experiments.</li></ul></ul>	675		Question: Does the paper provide open access to the data and code, with sufficient instruc-
<ul> <li>material?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/ public/guides/GodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https: //nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Guidelines:</li></ul></li></ul>	676		tions to faithfully reproduce the main experimental results, as described in supplemental
<ul> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including hor should provide instructions on data access and preparation, including hor the ave proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should revide scripts to reproduce all experimental results for thn new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should revide anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> </ul> </li> <li>Experimental Setting/Details         <ul> <li>Providing as much information as possible in supplemental material (appended to</li></ul></li></ul>	677		material?
<ul> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data, submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The anthors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If nolly a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> </ul> </li> <li>Guidelines:         <ul> <li>The authors the apper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experime</li></ul></li></ul>	678		Answer: [NA]
<ul> <li>Guidelines:</li> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/ public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https: //nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which news are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper- parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provide</li></ul>	679		Justification: The paper is a theory paper, and it has no experiments.
<ul> <li>The answer NA means that paper does not include experiments requiring code.</li> <li>Please see the NurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data are reproducible, they should state which ones are omitted from the script and why.</li> <li>A tsubmission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Guestion: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The full details scane be provided either with the code, in appendix, or as supplemental material.</li> <li>The full</li></ul>	680		Guidelines:
<ul> <li>Please see the NeurIPS code and data submission guidelines (https://nips.cc/ public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https: //nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper- parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The full details can be provide either with the</li></ul>	681		• The answer NA means that paper does not include experiments requiring code.
<ul> <li>public/guides/CodeSubmissionPolicy) for more details.</li> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nije.cc/public/Guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimental results for the new should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The experimental scripting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The experimental scripting should be presented in the core of the</li></ul>	682		• Please see the NeurIPS code and data submission guidelines (https://nips.cc/
<ul> <li>While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>6. Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>77 4 5</b></li></ul>	683		public/guides/CodeSubmissionPolicy) for more details.
<ul> <li>possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).</li> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The enswer NA means that the paper does not include experiments.</li> <li>The systemical Significance</li> <li>Question: Does the paper report pror bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li><b>The experimental Statistical Significance</b></li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li><b>The eathors Should answer</b> "Yes" if the results are acco</li></ul>	684		• While we encourage the release of code and data, we understand that this might not be
<ul> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https: //nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide sirplis to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental results for the new proposed method and baselines. If only a subset of experimental metrial (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Experimental Setting/Details</li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full de</li></ul>	685		possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is control to the contribution (a.g. for a new open source)
<ul> <li>The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https: //nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The daile is a new row and make sense of them.</li> <li>The full details can be provide either with the code, in appendix, or as supplemental material.</li> <li><b>Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance of the experiments?</li> <li>The full details can be provide either with the</li></ul>	686 687		herchmark)
<ul> <li>reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPiley) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimental results for the new should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>6. Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provide either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper is a theory paper, and it has no experiments.</li> <li>The dull details can be provide either with the code, in appendix, or as supplemental material.</li> <li>The dull details can be provide either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The austors s</li></ul></li></ul>	688		• The instructions should contain the exact command and environment needed to run to
<ul> <li>//nips.cc/public/guides/CodeSubmissionPolicy) for more details.</li> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Experimental Setting/Details</li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li>Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>The authors Should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance te</li></ul>	689		reproduce the results. See the NeurIPS code and data submission guidelines (https:
<ul> <li>The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>6. Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The full details can be provide either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>The answer: INA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The cult details can be provided either with the code, in appendix, or as supplemental material.</li> <li>The authors about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors about the statistical significance of the experiments.</li> <li>The authors about the statistical significance is a negret statistical significance is a provided either with the code, in appendix, or as supplemental material.</li> <li>The answer NA means</li></ul>	690		//nips.cc/public/guides/CodeSubmissionPolicy) for more details.
<ul> <li>to access the raw data, preprocessed data, intermediate data, and generated data, etc.</li> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:         <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li><b>Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li><b>Answer:</b> [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li><b>Cuestion:</b> Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li><b>Answer:</b> [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li><b>Cuestion:</b> Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the exp</li></ul>	691		• The authors should provide instructions on data access and preparation, including how
<ul> <li>The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experimenta are reproducible, they should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li><b>Guidelines:</b> <ul> <li>The answer NA means that the paper does not include experiments.</li> <li><b>Carperiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li><b>Caudelines:</b></li></ul></li></ul>	692		to access the raw data, preprocessed data, intermediate data, and generated data, etc.
<ul> <li>proposed method and baselines. If only a subset of experiments are reproducible, they</li> <li>should state which ones are omitted from the script and why.</li> <li>At submission time, to preserve anonymity, the authors should release anonymized</li> <li>versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the</li> <li>paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-</li> <li>parameters, how they were chosen, type of optimizer, etc.) necessary to understand the</li> <li>results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail</li> <li>that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental</li> <li>material.</li> <li><b>Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate</li> <li>information about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>The authors should aswer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eigen experiments)</li> </ul>	693		• The authors should provide scripts to reproduce all experimental results for the new
<ul> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper is a theory paper, and it has no experiments.</li> <li>The submity of the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The factors of xatistical significance tests, at least for the experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>Guidelines:</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing</li></ul>	694		proposed method and baselines. If only a subset of experiments are reproducible, they
<ul> <li>At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).</li> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li><b>6. Experimental Setting/Details</b></li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>71. Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>The answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by erro bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initializa</li></ul>	695		should state which ones are omitted from the script and why.
<ul> <li>Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.</li> <li>Experimental Setting/Details</li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li>7. Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines: <ul> <li>The answer INA</li> <li>Dass suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> </ul> </li> <li>The answer NA means that the paper does not include experiments.</li> <li>Guidelines: <ul> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eigen experimental conditions)</li> </ul></li></ul>	696 697		• At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable)
<ul> <li><sup>656</sup> paper) is recommended, but including URLs to data and code is permitted.</li> <li>6. Experimental Setting/Details</li> <li>Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li>7. Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines:</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall row</li> </ul>	608		<ul> <li>Providing as much information as possible in supplemental material (appended to the</li> </ul>
700       6. Experimental Setting/Details         701       Question: Does the paper specify all the training and test details (e.g., data splits, hyper- parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?         704       Answer: [NA]         705       Justification: The paper is a theory paper, and it has no experiments.         706       Guidelines:         707       • The answer NA means that the paper does not include experiments.         708       • The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.         710       • The full details can be provided either with the code, in appendix, or as supplemental material.         712       7. Experiment Statistical Significance         713       Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?         714       Justification: The paper is a theory paper, and it has no experiments.         715       Answer: [NA]         716       • The answer NA means that the paper does not include experiments.         715       Answer: [NA]         716       Justification: The paper is a theory paper, and it has no experiments.         717       Guidelines:         718       • The anthors should answer "Yes" if the results are ac	699		paper) is recommended, but including URLs to data and code is permitted.
701       Question: Does the paper specify all the training and test details (e.g., data splits, hyper-         702       parameters, how they were chosen, type of optimizer, etc.) necessary to understand the         703       results?         704       Answer: [NA]         705       Justification: The paper is a theory paper, and it has no experiments.         706       Guidelines:         707       • The answer NA means that the paper does not include experiments.         708       • The experimental setting should be presented in the core of the paper to a level of detail         709       • The full details can be provided either with the code, in appendix, or as supplemental         710       • The full details <b>Significance</b> 713       Question: Does the paper report error bars suitably and correctly defined or other appropriate         714       Instriction: The paper is a theory paper, and it has no experiments.         715       Answer: [NA]         716       Justification: The paper is a theory paper, and it has no experiments.         717       Guidelines:         718       • The answer NA means that the paper does not include experiments.         719       • The answer NA means that the paper does not include experiments.         719       • The answer NA means that the paper does not include experiments.         719       • The answer	700	6.	Experimental Setting/Details
<ul> <li>parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li>72 7. Experiment Statistical Significance <ul> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Guidelines: <ul> <li>The answer: [NA]</li> </ul> </li> <li>716 Justification: The paper is a theory paper, and it has no experiments.</li> <li>717 Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> </ul> </li> <li>718 • The answer NA means that the paper does not include experiments.</li> <li>719 • The answer NA means that the paper does not include experiments.</li> <li>719 • The answer NA means that the paper does not include experiments.</li> <li>719 • The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>728 • The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eiven experimental conditions)</li> </ul> </li> </ul>	701		Question: Does the paper specify all the training and test details (e.g., data splits, hyper-
<ul> <li>results?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li>7. Experiment Statistical Significance <ul> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments.</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eiven experimental conditions)</li> </ul> </li> </ul></li></ul>	702		parameters, how they were chosen, type of optimizer, etc.) necessary to understand the
<ul> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li>727 Experiment Statistical Significance <ul> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eigen experimental conditions)</li> </ul> </li> </ul></li></ul>	703		results?
<ul> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> </ul> </li> <li>727 Experiment Statistical Significance <ul> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul> </li> </ul></li></ul>	704		Answer: [NA]
706Guidelines:707• The answer NA means that the paper does not include experiments.708• The experimental setting should be presented in the core of the paper to a level of detail709that is necessary to appreciate the results and make sense of them.710• The full details can be provided either with the code, in appendix, or as supplemental711material.712 <b>7. Experiment Statistical Significance</b> 713Question: Does the paper report error bars suitably and correctly defined or other appropriate714information about the statistical significance of the experiments?715Answer: [NA]716Justification: The paper is a theory paper, and it has no experiments.717Guidelines:718• The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support729• The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eiven experimental conditions)	705		Justification: The paper is a theory paper, and it has no experiments.
<ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	706		Guidelines:
<ul> <li>The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with eiven experimental conditions)</li> </ul>	707		• The answer NA means that the paper does not include experiments.
<ul> <li>that is necessary to appreciate the results and make sense of them.</li> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li><b>7. Experiment Statistical Significance</b></li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	708		• The experimental setting should be presented in the core of the paper to a level of detail
<ul> <li>The full details can be provided either with the code, in appendix, or as supplemental material.</li> <li>7. Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	709		that is necessary to appreciate the results and make sense of them.
<ul> <li>material.</li> <li>7. Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul> </li> </ul>	710		• The full details can be provided either with the code, in appendix, or as supplemental
<ul> <li>7. Experiment Statistical Significance</li> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>7.15 Answer: [NA]</li> <li>7.16 Justification: The paper is a theory paper, and it has no experiments.</li> <li>7.17 Guidelines:</li> <li>7.18 • The answer NA means that the paper does not include experiments.</li> <li>7.19 • The authors should answer "Yes" if the results are accompanied by error bars, confi- dence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>7.18 • The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	711		material.
<ul> <li>Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul> </li> </ul>	712	7.	Experiment Statistical Significance
<ul> <li>information about the statistical significance of the experiments?</li> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines: <ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul> </li> </ul>	713		Question: Does the paper report error bars suitably and correctly defined or other appropriate
<ul> <li>Answer: [NA]</li> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	714		information about the statistical significance of the experiments?
<ul> <li>Justification: The paper is a theory paper, and it has no experiments.</li> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	715		Answer: [NA]
<ul> <li>Guidelines:</li> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	716		Justification: The paper is a theory paper, and it has no experiments.
<ul> <li>The answer NA means that the paper does not include experiments.</li> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	717		Guidelines:
<ul> <li>The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	718		• The answer NA means that the paper does not include experiments.
<ul> <li>dence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	719		• The authors should answer "Yes" if the results are accompanied by error bars, confi-
<ul> <li>the main claims of the paper.</li> <li>The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions)</li> </ul>	720		dence intervals, or statistical significance tests, at least for the experiments that support
<ul> <li>I he factors of variability that the error bars are capturing should be clearly stated (for</li> <li>example, train/test split, initialization, random drawing of some parameter, or overall</li> <li>run with given experimental conditions)</li> </ul>	721		the main claims of the paper.
run with given experimental conditions)	722		• I ne factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall
	724		run with given experimental conditions).

725 726		• The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
727		• The assumptions made should be given (e.g., Normally distributed errors).
728		• It should be clear whether the error bar is the standard deviation or the standard error
729		of the mean.
730		• It is OK to report 1-sigma error bars, but one should state it. The authors should
731		preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis
732		of Normality of errors is not verified.
733		• For asymmetric distributions, the authors should be careful not to show in tables or
734		figures symmetric error bars that would yield results that are out of range (e.g. negative
735		error rates).
736 737		• If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.
738	8.	Experiments Compute Resources
739		Question: For each experiment, does the paper provide sufficient information on the com-
740		puter resources (type of compute workers, memory, time of execution) needed to reproduce
741		the experiments?
742		Answer: [NA]
743		Justification: The paper is a theory paper, and it has no experiments.
744		Guidelines:
745		• The answer NA means that the paper does not include experiments.
746		• The paper should indicate the type of compute workers CPU or GPU, internal cluster,
747		or cloud provider, including relevant memory and storage.
748		• The paper should provide the amount of compute required for each of the individual
749		experimental runs as well as estimate the total compute.
750		• The paper should disclose whether the full research project required more compute
751 752		than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).
753	9.	Code Of Ethics
754		Question: Does the research conducted in the paper conform, in every respect, with the
755		NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?
756		Answer: [Yes]
757		Justification: The paper conforms with the NeurIPS Code of Ethics.
758		Guidelines:
759		• The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
760		• If the authors answer No, they should explain the special circumstances that require a
761		deviation from the Code of Ethics.
762		• The authors should make sure to preserve anonymity (e.g., if there is a special consid-
763		eration due to laws or regulations in their jurisdiction).
764	10.	Broader Impacts
765 766		Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?
767		Answer: [NA]
768		Justification: The paper provides a theoretical result. We do not believe there are potential
769		societal consequences of our work, aside from advancing the field of Machine Learning.
770		Guidelines:
771		• The answer NA means that there is no societal impact of the work performed.
772		• If the authors answer NA or No, they should explain why their work has no societal
773		impact or why the paper does not address societal impact.

774 775 776 777	• Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
778	• The conference expects that many papers will be foundational research and not tied
779	to particular applications, let alone deployments. However, if there is a direct path to
780	any negative applications, the authors should point it out. For example, it is legitimate
781	to point out that an improvement in the quality of generative models could be used to
782	generate deepfakes for disinformation. On the other hand, it is not needed to point out
783	that a generic algorithm for optimizing neural networks could enable people to train
784	models that generate Deepfakes faster.
785	• The authors should consider possible harms that could arise when the technology is
786	being used as intended and functioning correctly, harms that could arise when the
787	technology is being used as intended but gives incorrect results, and harms following
788	from (intentional or unintentional) misuse of the technology.
789	• If there are negative societal impacts, the authors could also discuss possible mitigation
790	strategies (e.g., gated release of models, providing defenses in addition to attacks,
791	mechanisms for monitoring misuse, mechanisms to monitor how a system learns from
792	feedback over time, improving the efficiency and accessibility of ML).
793 1	1. Safeguards
794	Question: Does the paper describe safeguards that have been put in place for responsible
795	release of data or models that have a high risk for misuse (e.g., pretrained language models,
796	image generators, or scraped datasets)?
797	Answer: [NA]
798	Justification: The paper only contains theoretical results.
799	Guidelines:
800	• The answer NA means that the paper poses no such risks.
801	• Released models that have a high risk for misuse or dual-use should be released with
802	necessary safeguards to allow for controlled use of the model, for example by requiring
803	that users adhere to usage guidelines or restrictions to access the model or implementing
804	safety filters.
805	• Datasets that have been scraped from the Internet could pose safety risks. The authors
806	should describe how they avoided releasing unsafe images.
807	• We recognize that providing effective safeguards is challenging, and many papers do
808	not require this, but we encourage authors to take this into account and make a best
809	faith effort.
810 1	2. Licenses for existing assets
811	Question: Are the creators or original owners of assets (e.g., code, data, models), used in
812	the paper, properly credited and are the license and terms of use explicitly mentioned and
813	properly respected?
814	Answer: [NA]
815	Justification: The paper only contains theoretical results. Previous work is properly cited.
816	Guidelines:
817	• The answer NA means that the paper does not use existing assets.
818	• The authors should cite the original paper that produced the code package or dataset.
910	• The authors should state which version of the asset is used and if possible include a
820	URL.
821	• The name of the license (e.g., CC-BY 4.0) should be included for each asset.
822	• For scraped data from a particular source (e.g., website), the copyright and terms of
823	service of that source should be provided.
824	• If assets are released, the license, copyright information, and terms of use in the
825	package should be provided. For popular datasets, paperswithcode.com/datasets
826	has curated licenses for some datasets. Their licensing guide can help determine the
827	license of a dataset.

828 829		• For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
830		• If this information is not available online, the authors are encouraged to reach out to
831		the asset's creators.
832	13.	New Assets
833 834		Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?
835		Answer: [NA]
836		Justification: The paper does not introduce any new asset.
837		Guidelines:
838		• The answer NA means that the paper does not release new assets
839		<ul> <li>Researchers should communicate the details of the dataset/code/model as part of their</li> </ul>
840 841		submissions via structured templates. This includes details about training, license, limitations, etc.
842		• The paper should discuss whether and how consent was obtained from people whose
843		asset is used.
844 845		• At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.
946	14	Crowdsourcing and Research with Human Subjects
040	17.	Ousstian, East answed surveying a surveying and reasonable with human subjects
847		include the full text of instructions given to participants and screenshots, if applicable as
849		well as details about compensation (if any)?
850		Answer: [NA]
851		Justification: This point does not apply to our paper
950		Guidelines:
852		The NA was determined in the second s
853 854		• The answer NA means that the paper does not involve crowdsourcing nor research with human subjects
955		<ul> <li>Including this information in the supplemental material is fine, but if the main contribu-</li> </ul>
856		tion of the paper involves human subjects, then as much detail as possible should be
857		included in the main paper.
858		• According to the NeurIPS Code of Ethics, workers involved in data collection, curation,
859		or other labor should be paid at least the minimum wage in the country of the data
860		collector.
861	15.	Institutional Review Board (IRB) Approvals or Equivalent for Research with Human
862		Subjects
863		Question: Does the paper describe potential risks incurred by study participants, whether
864		such risks were disclosed to the subjects, and whether Institutional Review Board (IRB)
865 866		institution) were obtained?
967		Answer: [NA]
007		Justification: This point does not apply to our paper
000		Guidelines:
869		The NA was determined in the second s
870		• The answer NA means that the paper does not involve crowdsourcing nor research with human subjects
970		<ul> <li>Depending on the country in which research is conducted IRR approval (or equivalent).</li> </ul>
873		may be required for any human subjects research. If you obtained IRB approval, you
874		should clearly state this in the paper.
875		• We recognize that the procedures for this may vary significantly between institutions
876		and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the
877		guidelines for their institution.
878 879		• For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.