Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems

Yuxin Chen * Emmanuel J. Candès *†

May 2015; Revised January 2016

Abstract

We consider the fundamental problem of solving quadratic systems of equations in n variables, where $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2$, $i = 1, \ldots, m$ and $\boldsymbol{x} \in \mathbb{R}^n$ is unknown. We propose a novel method, which starting with an initial guess computed by means of a spectral method, proceeds by minimizing a nonconvex functional as in the Wirtinger flow approach [1]. There are several key distinguishing features, most notably, a distinct objective functional and novel update rules, which operate in an adaptive fashion and drop terms bearing too much influence on the search direction. These careful selection rules provide a tighter initial guess, better descent directions, and thus enhanced practical performance. On the theoretical side, we prove that for certain unstructured models of quadratic systems, our algorithms return the correct solution in linear time, i.e. in time proportional to reading the data $\{\boldsymbol{a}_i\}$ and $\{y_i\}$ as soon as the ratio m/n between the number of equations and unknowns exceeds a fixed numerical constant. We extend the theory to deal with noisy systems in which we only have $y_i \approx |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2$ and prove that our algorithms achieve a statistical accuracy, which is nearly un-improvable. We complement our theoretical study with numerical examples showing that solving random quadratic systems is both computationally and statistically not much harder than solving linear systems of the same size—hence the title of this paper. For instance, we demonstrate empirically that the computational cost of our algorithm is about four times that of solving a least-squares problem of the same size.

1 Introduction

1.1 Problem formulation

Imagine we are given a set of m quadratic equations taking the form

$$y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2, \qquad i = 1, \cdots, m, \tag{1}$$

where the data $\boldsymbol{y} = [y_i]_{1 \le i \le m}$ and design vectors $\boldsymbol{a}_i \in \mathbb{R}^n / \mathbb{C}^n$ are known whereas $\boldsymbol{x} \in \mathbb{R}^n / \mathbb{C}^n$ is unknown. Having information about $|\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2$—or, equivalently, $|\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|$—means that we a priori know nothing about the phases or signs of the linear products $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle$. The problem is this: can we hope to identify a solution, if any, compatible with this nonlinear system of equations?

This problem is combinatorial in nature as one can alternatively pose it as recovering the missing signs of $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle$ from magnitude-only observations. As is well known, many classical combinatorial problems with Boolean variables may be cast as special instances of (1). As an example, consider the NP-hard *stone problem* [2] in which we have n stones each of weight $w_i > 0$ ($1 \le i \le n$), which we would like to divide into two groups of equal sum weight. Letting $x_i \in \{-1, 1\}$ indicate which of the two groups the ith stone belongs to, one can formulate this problem as solving the following quadratic system

$$\begin{cases} x_i^2 = 1, & i = 1, \cdots, n, \\ (w_1 x_1 + \cdots + w_n x_n)^2 = 0. \end{cases} \tag{2}$$

*Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.
†Department of Mathematics, Stanford University, Stanford, CA 94305, U.S.A.

However simple this formulation may seem, even checking whether a solution to (2) exists or not is known to be NP-hard.

Moving from combinatorial optimization to the physical sciences, one application of paramount importance is the *phase retrieval* [3,4] problem, which permeates through a wide spectrum of techniques including X-ray crystallography, diffraction imaging, microscopy, and even quantum mechanics. In a nutshell, the problem of phase retrieval arises due to the physical limitation of optical sensors, which are often only able to record the intensities of the diffracted waves scattered by an object under study. Notably, upon illuminating an object $\boldsymbol{x}$, the diffraction pattern is of the form of $\boldsymbol{A}\boldsymbol{x}$; however, it is only possible to obtain intensity measurements $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$ leading to the quadratic system (1).[1] In the Fraunhofer regime where data is collected in the far-field zone, $\boldsymbol{A}$ is given by the spatial Fourier transform. We refer to [5] for in-depth reviews of this subject.

Continuing this motivating line of thought, in any real-world application recorded intensities are always corrupted by at least a small amount of noise so that observed data are only about $|\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2$; i.e.

$$y_i \approx |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2, \qquad i = 1, \cdots, m. \tag{3}$$

Although we present results for arbitrary noise distributions—even for non-stochastic noise—we shall pay particular attention to the Poisson data model, which assumes

$$y_i \overset{\text{ind.}}{\sim} \mathsf{Poisson}\big(\, |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2 \,\big), \qquad i = 1, \cdots, m. \tag{4}$$

The reason why this statistical model is of special interest is that it naturally describes the variation in the number of photons detected by an optical sensor in various imaging applications.

1.2 Nonconvex optimization

Under a stochastic noise model with independent samples, a first impulse for solving (3) is to seek the maximum likelihood estimate (MLE), namely,

$$\text{minimize}_{\boldsymbol{z}} \quad -\sum\nolimits_{i=1}^{m} \ell\left(\boldsymbol{z}; y_i\right), \tag{5}$$

where $\ell\left(\boldsymbol{z}; y_i\right)$ denotes the log-likelihood of a candidate solution $\boldsymbol{z}$ given the outcome y_i. For instance, under the Poisson data model (4) one can write

$$\ell(\boldsymbol{z}; y_i) = y_i \log(|\boldsymbol{a}_i^* \boldsymbol{z}|^2) - |\boldsymbol{a}_i^* \boldsymbol{z}|^2 \tag{6}$$

modulo some constant offset. Unfortunately, the log-likelihood is usually not concave, thus making the problem of computing the MLE NP-hard in general.

To alleviate this computational intractability, several convex surrogates have been proposed that work particularly well when the design vectors $\{\boldsymbol{a}_i\}$ are chosen at random [6–20]. The basic idea is to introduce a rank-one matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^*$ to linearize the quadratic constraints, and then relax the rank-one constraint. Suppose we have Poisson data, then this strategy converts the problem into a convenient convex program:

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{X}} & \sum_{i=1}^{m} (\mu_i - y_i \log \mu_i) + \lambda \mathsf{Tr}(\boldsymbol{X}) \\ \text{subject to} & \mu_i = \boldsymbol{a}_i^\top \boldsymbol{X} \boldsymbol{a}_i, \quad 1 \le i \le m, \\ & \boldsymbol{X} \succeq \boldsymbol{0}. \end{array}$$

Note that the log-likelihood function is augmented by the trace functional $\mathsf{Tr}(\cdot)$ whose role is to promote low-rank solutions. While such convex relaxation schemes enjoy intriguing performance guarantees in many aspects (e.g. they achieve minimal sample complexity and near-optimal error bounds for certain noise models), the computational cost typically far exceeds the order of n^3. This limits applicability to large-dimensional data.

This paper follows another route: rather than lifting the problem into higher dimensions by introducing matrix variables, this paradigm maintains its iterates within the vector domain and optimize the nonconvex

[1] Here and below, for $\boldsymbol{z} \in \mathbb{C}^n$, $|\boldsymbol{z}|$ (resp. $|\boldsymbol{z}|^2$) represents the vector of magnitudes $(|z_1|, \cdots, |z_n|)^\top$ (resp. squared magnitudes $(|z_1|^2, \cdots, |z_n|^2)^\top$).

objective directly (e.g. [1, 3, 21–28]). One promising approach along this line is the recently proposed two-stage algorithm called *Wirtinger Flow* (WF) [1]. Simply put, WF starts by computing a suitable initial guess $\boldsymbol{z}^{(0)}$ using a spectral method, and then successively refines the estimate via an update rule that bears a strong resemblance to a gradient descent scheme, namely,

$$\boldsymbol{z}^{(t+1)} = \boldsymbol{z}^{(t)} + \frac{\mu_t}{m} \sum\nolimits_{i=1}^{m} \nabla \ell(\boldsymbol{z}^{(t)}; y_i),$$

where $\boldsymbol{z}^{(t)}$ denotes the tth iterate of the algorithm, and μ_t is the step size (or learning rate). Here, $\nabla \ell(\boldsymbol{z}; y_i)$ stands for the Wirtinger derivative w.r.t. $\boldsymbol{z}$, which in the real-valued case reduces to the ordinary gradient. The main results of [1] demonstrate that WF is surprisingly accurate for independent Gaussian design. Specifically, when $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ or $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) + j\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$:

1. WF achieves exact recovery from $m = O(n \log n)$ quadratic equations when there is no noise;[2]
2. WF attains ϵ-accuracy—in a relative sense—within $O(mn^2 \log(1/\epsilon))$ time (or flops);
3. In the presence of Gaussian noise, WF is stable and converges to the MLE as shown in [29].

While these results formalize the advantages of WF, the computational complexity of WF is still much larger than the best one can hope for. Moreover, the statistical guarantee in terms of the sample complexity is weaker than that achievable by convex relaxations.[3]

1.3 This paper: Truncated Wirtinger Flow

This paper develops an efficient linear-time algorithm, which also enjoys near-optimal statistical guarantees. Following the spirit of WF, we propose a novel procedure called *Truncated Wirtinger Flow* (TWF) adopting a more adaptive gradient flow. Informally, TWF proceeds in two stages:

1. **Initialization:** compute an initial guess $\boldsymbol{z}^{(0)}$ by means of a spectral method applied to a subset $\mathcal{T}_0$ of the observations $\{y_i\}$;
2. **Loop:** for $0 \le t < T$,
$$\boldsymbol{z}^{(t+1)} = \boldsymbol{z}^{(t)} + \frac{\mu_t}{m} \sum_{i \in \mathcal{T}_{t+1}} \nabla \ell(\boldsymbol{z}^{(t)}; y_i) \tag{7}$$
for some index subset $\mathcal{T}_{t+1} \subseteq \{1, \cdots, m\}$ determined by $\boldsymbol{z}^{(t)}$.

Three remarks are in order.

- Firstly, we regularize both the initialization and the gradient flow in a data-dependent fashion by operating only upon some iteration-varying index subsets $\mathcal{T}_t$. This is a distinguishing feature of TWF in comparison to WF and other gradient descent variants. In words, $\mathcal{T}_t$ corresponds to those data $\{y_i\}$ whose resulting spectral or gradient components are in some sense not excessively large; see Sections 2 and 3 for details. As we shall see later, the main point is that this careful data trimming procedure gives us a tighter initial guess and more stable search directions.
- Secondly, we recommend that the step size μ_t is either taken as some appropriate constant or determined by a backtracking line search. For instance, under appropriate conditions, we can take $\mu_t = 0.2$ for all t.

[2] The standard notation $f(n) = O(g(n))$ or $f(n) \lesssim g(n)$ (resp. $f(n) = \Omega(g(n))$ or $f(n) \gtrsim g(n)$) means that there exists a constant $c > 0$ such that $|f(n)| \le c|g(n)|$ (resp. $|f(n)| \ge c|g(n)|$). $f(n) \asymp g(n)$ means that there exist constants $c_1, c_2 > 0$ such that $c_1|g(n)| \le |f(n)| \le c_2|g(n)|$.

[3] M. Soltanolkotabi recently informed us that the sample complexity of WF may be improved if one employs a better initialization procedure.

- Finally, the most expensive part of the gradient stage consists in computing $\nabla \ell(\boldsymbol{z}; y_i)$, $1 \le i \le m$, which can often be performed in an efficient manner. More concretely, under the *real-valued* Poisson data model (4) one has

$$\nabla \ell(\boldsymbol{z}; y_i) = 2 \left\{ \frac{y_i}{|\boldsymbol{a}_i^\top \boldsymbol{z}|^2} \boldsymbol{a}_i \boldsymbol{a}_i^\top \boldsymbol{z} - \boldsymbol{a}_i \boldsymbol{a}_i^\top \boldsymbol{z} \right\} = 2 \left(\frac{y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \right) \boldsymbol{a}_i,$$

  Thus, calculating $\{\nabla \ell(\boldsymbol{z}; y_i)\}$ essentially amounts to two matrix-vector products. Letting $\boldsymbol{A} := [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m]^\top$ as before, we have

$$\sum_{i \in \mathcal{T}_{t+1}} \nabla \ell(\boldsymbol{z}^{(t)}; y_i) = \boldsymbol{A}^\top \boldsymbol{v}, \qquad v_i = \begin{cases} 2 \frac{y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}, & i \in \mathcal{T}_{t+1}, \\ 0, & \text{otherwise.} \end{cases}$$

  Hence, $\boldsymbol{A}\boldsymbol{z}$ gives $\boldsymbol{v}$ and $\boldsymbol{A}^\top \boldsymbol{v}$ the desired regularized gradient.

A detailed specification of the algorithm is deferred to Section 2.

1.4 Numerical surprises

To give the readers a sense of the practical power of TWF, we present here three illustrative numerical examples. Since it is impossible to recover the global sign—i.e. we cannot distinguish $\boldsymbol{x}$ from $-\boldsymbol{x}$—we will evaluate our solutions to the quadratic equations through the distance measure put forth in [1] representing the Euclidean distance modulo a global sign: for complex-valued signals,

$$\operatorname{dist}(\boldsymbol{z}, \boldsymbol{x}) := \min_{\varphi:\in[0,2\pi)} \|e^{-j\varphi} \boldsymbol{z} - \boldsymbol{x}\|, \tag{8}$$

while it is simply $\min \|\boldsymbol{z} \pm \boldsymbol{x}\|$ in the real-valued case. We shall use $\operatorname{dist}(\hat{\boldsymbol{x}}, \boldsymbol{x})/\|\boldsymbol{x}\|$ throughout to denote the relative error of an estimate $\hat{\boldsymbol{x}}$. In the sequel, TWF proceeds by attempting to maximize the Poisson log-likelihood (6). Standalone Matlab implementations of TWF are available at `http://statweb.stanford.edu/~candes/publications.html` (see [30] for straight WF implementations).

The first numerical example concerns the following two problems under noiseless real-valued data:

$$\begin{array}{lllll} \text{(a)} & \text{find } \boldsymbol{x} \in \mathbb{R}^n & \text{s.t. } b_i = \boldsymbol{a}_i^\top \boldsymbol{x}, & & 1 \le i \le m; \\ \text{(b)} & \text{find } \boldsymbol{x} \in \mathbb{R}^n & \text{s.t. } b_i = |\boldsymbol{a}_i^\top \boldsymbol{x}|, & & 1 \le i \le m. \end{array}$$

Apparently, (a) involves solving a linear system of equations (or a linear least squares problem), while (b) is tantamount to solving a quadratic system. Arguably the most popular method for solving large-scale least squares problems is the conjugate gradient (CG) method [31] applied to the normal equations. We are going to compare the computational efficiency between CG (for solving least squares) and TWF with a step size $\mu_t \equiv 0.2$ (for solving a quadratic system). Set $m = 8n$ and generate $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $1 \le i \le m$, independently. This gives a matrix $\boldsymbol{A}^\top \boldsymbol{A}$ with a low condition number equal to about $(1 + \sqrt{1/8})^2/(1 - \sqrt{1/8})^2 \approx 4.38$ by the Marchenko-Pastur law. Therefore, this is an ideal setting for CG as it converges extremely rapidly [32, Theorem 38.5]. Fig. 1 shows the relative estimation error of each method as a function of the iteration count, where TWF is seeded through 10 power iterations. For ease of comparison, we illustrate the iteration counts in different scales so that 4 TWF iterations are equivalent to 1 CG iteration.

Recognizing that each iteration of CG and TWF involves two matrix vector products $\boldsymbol{A}\boldsymbol{z}$ and $\boldsymbol{A}^\top \boldsymbol{v}$, for such a design we reach a suprising observation:

> *Even when all phase information is missing, TWF is capable of solving a quadratic system of equations only about 4 times slower than solving a least squares problem of the same size!*

To illustrate the applicability of TWF on real images, we turn to testing our algorithm on a digital photograph of Stanford main quad containing 320×1280 pixels. We consider a type of measurements that falls under the category of coded diffraction patterns (CDP) [33] and set

$$\boldsymbol{y}^{(l)} = |\boldsymbol{F}\boldsymbol{D}^{(l)}\boldsymbol{x}|^2, \qquad 1 \le l \le L. \tag{9}$$

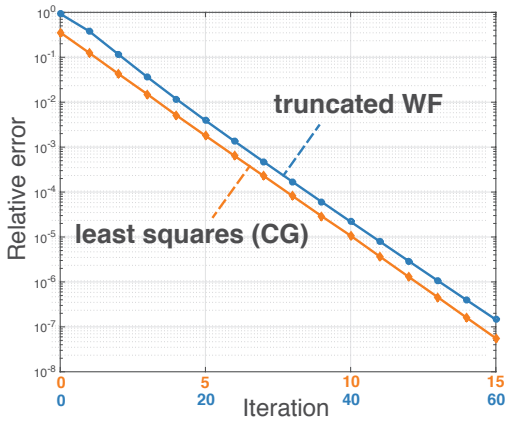


Figure 1: Relative errors of CG and TWF vs. iteration count. Here, $n = 1000$, $m = 8n$, and TWF is seeded using just 10 power iterations.

Here, $\boldsymbol{F}$ stands for a discrete Fourier transform (DFT) matrix, and $\boldsymbol{D}^{(l)}$ is a diagonal matrix whose diagonal entries are independently and uniformly drawn from $\{1, -1, j, -j\}$ (phase delays). In phase retrieval, each $\boldsymbol{D}^{(l)}$ represents a random mask placed after the object so as to modulate the illumination patterns. When L masks are employed, the total number of quadratic measurements is $m = nL$. In this example, $L = 12$ random coded patterns are generated to measure each color band (i.e. red, green, or blue) separately. The experiment is carried out on a MacBook Pro equipped with a 3 GHz Intel Core i7 and 16GB of memory. We run 50 iterations of the truncated power method for initialization, and 50 regularized gradient iterations, which in total costs 43.9 seconds or 2400 FFTs for each color band. The relative error after regularized spectral initialization and after 50 TWF iterations are 0.4773 and 2.16×10^{-5}, respectively, with the recovered images displayed in Fig. 2. In comparison, the spectral initialization using 50 untruncated power iterations returns an image of relative error 1.409, which is almost like a random guess and extremely far from the truth.

While the above experiments concern noiseless data, the numerical surprise extends to the noisy realm. Suppose the data are drawn according to the Poisson noise model (4), with $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ independently generated. Fig. 3 displays the empirical relative mean-square error (MSE) of TWF as a function of the signal-to-noise ratio (SNR), where the relative MSE for an estimate $\hat{\boldsymbol{x}}$ and the SNR are defined as[4]

$$\text{MSE} := \frac{\text{dist}^2(\hat{\boldsymbol{x}}, \boldsymbol{x})}{\|\boldsymbol{x}\|^2}, \quad \text{and} \quad \text{SNR} := 3\|\boldsymbol{x}\|^2. \tag{10}$$

Both SNR and MSE are displayed on a dB scale (i.e. the values of $10\log_{10}(\text{SNR})$ and $10\log_{10}(\text{rel. MSE})$ are plotted). To evaluate the accuracy of the TWF solutions, we consider the performance achieved by MLE applied to an *ideal* problem in which the true phases are revealed. In this ideal scenario, in addition to the data $\{y_i\}$ we are further given exact phase information $\{\varphi_i = \text{sign}(\boldsymbol{a}_i^\top \boldsymbol{x})\}$. Such precious information gives away the phase retrieval problem and makes the MLE efficiently solvable since the MLE problem with side information

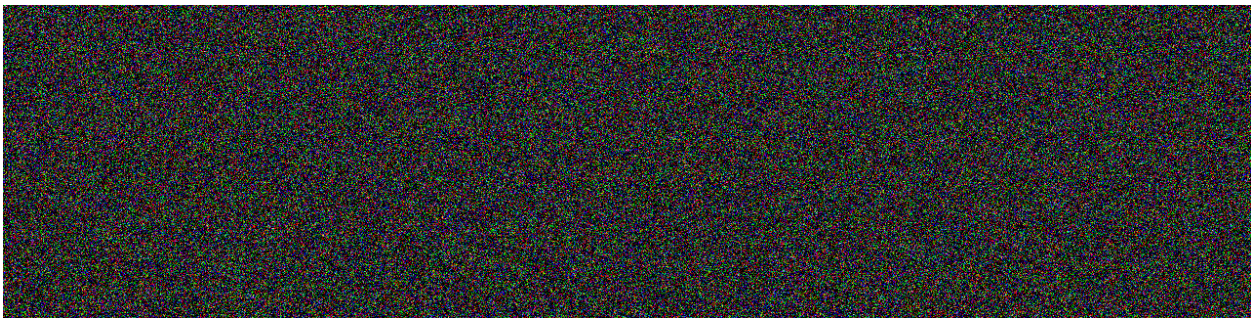
$$\begin{array}{ll} \text{minimize}_{\boldsymbol{z}\in\mathbb{R}^n} & -\sum_{i=1}^m y_i \log\left(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right) + (\boldsymbol{a}_i^\top \boldsymbol{z})^2 \\ \text{subject to} & \varphi_i = \text{sign}(\boldsymbol{a}_i^\top \boldsymbol{z}) \end{array}$$

can be cast as a convex program

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^n} \quad -\sum_{i=1}^m 2y_i \log\left(\varphi_i \boldsymbol{a}_i^\top \boldsymbol{z}\right) + (\boldsymbol{a}_i^\top \boldsymbol{z})^2.$$

Fig. 3 illustrates the empirical performance for this ideal problem. The plots demonstrate that even when all phases are erased, TWF yields a solution of nearly the best possible quality, since it only incurs an extra 1.5

[4]To justify the definition of SNR, note that the signals and noise are captured by $\mu_i = (\boldsymbol{a}_i^\top \boldsymbol{x})^2$ and $y_i - \mu_i$, $1 \le i \le m$, respectively. The ratio of the signal power to the noise power is therefore $\frac{\sum_{i=1}^m \mu_i^2}{\sum_{i=1}^m \mathbf{Var}[y_i]} = \frac{\sum_{i=1}^m |\boldsymbol{a}_i^\top \boldsymbol{x}|^4}{\sum_{i=1}^m |\boldsymbol{a}_i^\top \boldsymbol{x}|^2} \approx \frac{3m\|\boldsymbol{x}\|^4}{m\|\boldsymbol{x}\|^2} = 3\|\boldsymbol{x}\|^2$.



(a)

(b)

(c)

Figure 2: The recovered images after (a) spectral initialization; (b) regularized spectral initialization; and (c) 50 TWF gradient iterations following the regularized initiliazation.

dB loss compared to ideal MLE computed with all true phases revealed. This phenomenon arises regardless of the SNR!

1.5 Main results

The preceding numerical discoveries unveil promising features of TWF in three aspects: (1) exponentially fast convergence; (2) exact recovery from noiseless data with sample complexity $O(n)$; (3) nearly minimal mean-square loss in the presence of noise. This paper offers a formal substantiation of all these findings. To this end, we assume a tractable model in which the design vectors $\boldsymbol{a}_i$'s are independent Gaussian:

$$\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n). \tag{11}$$

For concreteness, our results are concerned with TWF designed based on the Poisson log-likelihood function

$$\ell_i(\boldsymbol{z}) := \ell(\boldsymbol{z}; y_i) = y_i \log\left(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right) - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2, \tag{12}$$

where we shall use $\ell_i(\boldsymbol{z})$ as a shorthand for $\ell(\boldsymbol{z}; y_i)$ from now on. We begin with the performance guarantees of TWF in the absence of noise.

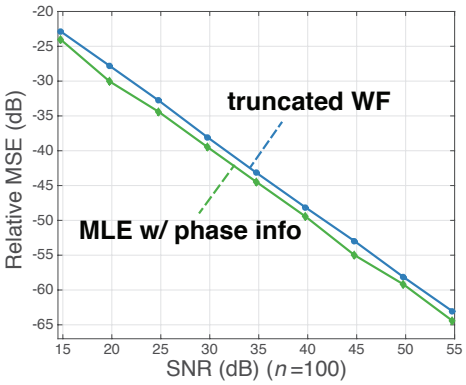


Figure 3: Relative MSE vs. SNR in dB. The curves are shown for two settings: TWF for solving quadratic equations (blue), and MLE had we observed additional phase information (green). The results are shown for $n = 100$, and each point is averaged over 50 Monte Carlo trials.

Theorem 1 (**Exact recovery**). *Consider the noiseless case (1) with an arbitrary signal $\boldsymbol{x} \in \mathbb{R}^n$. Suppose that the step size μ_t is either taken to be a positive constant $\mu_t \equiv \mu$ or chosen via a backtracking line search. Then there exist some universal constants $0 < \rho, \nu < 1$ and $\mu_0, c_0, c_1, c_2 > 0$ such that with probability exceeding $1 - c_1 \exp(-c_2 m)$, the truncated Wirtinger Flow estimates (Algorithm 1 with parameters specified in Table 1) obey*

$$\operatorname{dist}(\boldsymbol{z}^{(t)}, \boldsymbol{x}) \quad \leq \quad \nu(1-\rho)^t \|\boldsymbol{x}\|, \quad \forall t \in \mathbb{N}, \tag{13}$$

provided that

$$m \geq c_0 n \quad \text{and} \quad 0 < \mu \leq \mu_0.$$

As explained below, we can often take $\mu_0 \approx 0.3$.

Remark 1. As will be made precise in Section 5 (and in particular Proposition 1), one can take

$$\mu_0 = \frac{0.994 - \zeta_1 - \zeta_2 - \sqrt{2/(9\pi)}\alpha_h^{-1}}{2\left(1.02 + 0.665/\alpha_h\right)}$$

for some small quantities ζ_1, ζ_2 and some predetermined threshold α_h that is usually taken to be $\alpha_h \geq 5$. Under appropriate conditions, one can treat μ_0 as $\mu_0 \approx 0.3$.

Theorem 1 justifies at least two appealing features of TWF: (i) *minimal sample complexity* and (ii) *linear-time computational cost*. Specifically, TWF allows exact recovery from $O(n)$ quadratic equations, which is optimal since one needs at least n measurements to have a well-posed problem. Also, because of the geometric convergence rate, TWF achieves ϵ-accuracy (i.e. $\operatorname{dist}(\boldsymbol{z}^{(t)}, \boldsymbol{x}) \leq \epsilon \|\boldsymbol{x}\|$) within at most $O(\log(1/\epsilon))$ iterations. The total computational cost is therefore $O(mn \log(1/\epsilon))$, which is linear in the problem size. These outperform the performance guarantees of WF [1], which runs in $O(mn^2 \log(1/\epsilon))$ time and requires $O(n \log n)$ sample complexity.

We emphasize that enhanced performance vis-à-vis WF is not the result of a sharper analysis, but rather, the result of key algorithmic changes. In both the initialization and iterative refinement stages, TWF proceeds in a more prudent manner by means of proper regularization, which effectively trims away those components that are too influential on either the initial guess or search directions, thus reducing the volatility of each movement. With a tighter initialization and better-controlled search directions in place, we take the step size in a far more liberal fashion—which is some constant bounded away from 0—compared to a step size which is $O(1/n)$ as explained in [1]. In fact, what enables the movement to be more aggressive is exactly the cautious choice of $\mathcal{T}_t$, which precludes adverse effects from high-leverage samples.

To be broadly applicable, the proposed algorithm must guarantee reasonably faithful estimates in the presence of noise. Suppose that

$$y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2 + \eta_i, \qquad 1 \leq i \leq m, \tag{14}$$

where η_i represents an error term. We claim that TWF is stable against additive noise, as demonstrated in the theorem below.

Theorem 2 (**Stability**). *Consider the noisy case (14). Suppose that the step size μ_t is either taken to be a positive constant $\mu_t \equiv \mu$ or chosen via a backtracking line search. If*

$$m \geq c_0 n, \quad \mu \leq \mu_0, \quad \text{and} \quad \|\boldsymbol{\eta}\|_\infty \leq c_1 \|\boldsymbol{x}\|^2, \tag{15}$$

then with probability at least $1 - c_2 \exp(-c_3 m)$, the truncated Wirtinger Flow estimates (Algorithm 1 with parameters specified in Table 1) satisfy

$$\operatorname{dist}(\boldsymbol{z}^{(t)}, \boldsymbol{x}) \lesssim \frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{x}\|} + (1-\rho)^t \|\boldsymbol{x}\|, \quad \forall t \in \mathbb{N} \tag{16}$$

simultanesouly for all $\boldsymbol{x} \in \mathbb{R}^n$. Here, $0 < \rho < 1$ and $\mu_0, c_0, c_1, c_2, c_3 > 0$ are some universal constants.

Under the Poisson noise model (4), one has

$$\operatorname{dist}(\boldsymbol{z}^{(t)}, \boldsymbol{x}) \lesssim 1 + (1-\rho)^t \|\boldsymbol{x}\|, \quad \forall t \in \mathbb{N} \tag{17}$$

with probability approaching one, provided that $\|\boldsymbol{x}\| \geq \log^{1.5} m$.

Remark 2. In the main text, we will prove Theorem 2 only for the case where $\boldsymbol{x}$ is fixed and independent of the design vectors $\{\boldsymbol{a}_i\}$. Interested readers are referred to the supplemental materials [34] for the proof of the universal theory (i.e. the case simultaneously accommodating all $\boldsymbol{x} \in \mathbb{R}^n$). Note that when there is no noise ($\boldsymbol{\eta} = \boldsymbol{0}$), this stronger result guarantees the universality of the noiseless recovery.

Remark 3. [29] establishes stability estimates using the WF approach under Gaussian noise. There, the sample and computational complexities are still on the order of $n \log n$ and mn^2 respectively whereas the computational complexity in Theorem 2 is linear, i.e. on the order of mn.

Theorem 2 essentially reveals that the estimation error of TWF rapidly shrinks to $O\left(\frac{\|\boldsymbol{\eta}\|/\sqrt{m}}{\|\boldsymbol{x}\|}\right)$ within logarithmic iterations. Put another way, since the SNR for the model (14) is captured by

$$\mathrm{SNR} := \frac{\sum_{i=1}^m |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^4}{\|\boldsymbol{\eta}\|^2} \approx \frac{3m\|\boldsymbol{x}\|^4}{\|\boldsymbol{\eta}\|^2}, \tag{18}$$

we immediately arrive at an alternative form of the performance guarantee:

$$\operatorname{dist}(\boldsymbol{z}^{(t)}, \boldsymbol{x}) \lesssim \frac{1}{\sqrt{\mathrm{SNR}}}\|\boldsymbol{x}\| + (1-\rho)^t \|\boldsymbol{x}\|, \quad \forall t \in \mathbb{N}, \tag{19}$$

revealing the stability of TWF as a function of SNR. We emphasize that this estimate holds for any error term $\boldsymbol{\eta}$—i.e. any noise structure, even deterministic. This being said, specializing this estimate to the Poisson noise model (4) with $\|\boldsymbol{x}\| \gtrsim \log^{1.5} m$ gives an estimation error that will eventually approach a numerical constant, independent of n and m.

Encouragingly, this is already the best statistical guarantee any algorithm can achieve. We formalize this claim by deriving a fundamental lower bound on the minimax estimation error.

Theorem 3 (**Lower bound on the minimax risk**). *Suppose that $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $m = \kappa n$ for some fixed κ independent of n, and n is sufficiently large. For any $K \geq \log^{1.5} m$, define*[5]

$$\Upsilon(K) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \in (1 \pm 0.1)K\}.$$

With probability approaching one, the minimax risk under the Poisson model (4) obeys

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x} \in \Upsilon(K)} \mathbb{E}\big[\operatorname{dist}(\hat{\boldsymbol{x}}, \boldsymbol{x}) \;\big|\; \{\boldsymbol{a}_i\}_{1 \leq i \leq m}\big] \geq \frac{\varepsilon_1}{\sqrt{\kappa}}, \tag{20}$$

where the infimum is over all estimator $\hat{\boldsymbol{x}}$. Here, $\varepsilon_1 > 0$ is a numerical constant independent of n and m.

[5]Here, 0.1 can be replaced by any positive constant within (0, 1/2).

When the number m of measurements is proportional to n and the energy of the planted solution exceeds $\log^3 m$, Theorem 3 asserts that there exists absolutely no estimator that can achieve an estimation error that vanishes as n increases. This lower limit matches the estimation error of TWF, which corroborates the optimality of TWF under noisy data.

Recall that in many optical imaging applications, the output data we collect are the intensities of the diffractive waves scattered by the sample or specimen under study. The Poisson noise model employs the input $\boldsymbol{x}$ and output $\boldsymbol{y}$ to describe the numbers of photons diffracted by the specimen and detected by the optical sensor, respectively. Each specimen needs to be sufficiently illuminated in order for the receiver to sense the diffracted light. In such settings, the low-intensity regime $\|\boldsymbol{x}\| \le \log^{1.5} m$ is of little practical interest as it corresponds to an illumination with just very few photons. We forego the details.

It is worth noting that apart from WF, various other nonconvex procedures have been proposed as well for phase retrieval, including the error reduction schemes dating back to Gerchberg-Saxton and Fienup [3, 4], iterated projections [22], alternating minimization [21], generalized approximate message passing [23], Kaczmarz method [35], and greedy methods that exploit additional sparsity constraint [27], to name just a few. While these paradigms enjoy favorable empirical behavior, most of them fall short of theoretical support, except for a version of alternating minimization (called AltMinPhase) [21] that requires fresh samples for each iteration. In comparison, AltMinPhase attains ϵ-accuracy when the sample complexity exceeds the order of $n\log^3 n + n\log^2 n \log(1/\epsilon)$, which is at least a factor of $\log^3 n$ from optimal and is empirically largely outperformed by the variant that reuses all samples. In contrast, our algorithm uses the same set of samples all the time and is therefore practically appealing. Furthermore, none of these algorithms come with provable stability guarantees, which are particularly important in most realistic scenarios. Numerically, each iteration of Fienup's algorithm (or alternating minimization) involves solving a least squares problem, and the algorithm converges in tens or hundreds of iterations. This is computationally more expensive than TWF, whose computational complexity is merely about 4 times that of solving a least squares problem. Interesting readers are referred to [1] for a comparison of several non-convex schemes, and [33] for a discussion of other alternative approaches (e.g. [36, 37]) and performance lower bounds (e.g. [38, 39]).

2 Algorithm: Truncated Wirtinger Flow

This section describes the two stages of TWF in details, presented in a reverse order. For each stage, we start with some algorithmic issues encountered by WF, which is then used to motivate and explain the basic principles of TWF. Here and throughout, we let $\mathcal{A}: \mathbb{R}^{n\times n} \mapsto \mathbb{R}^m$ be the linear map

$$\boldsymbol{M} \in \mathbb{R}^{n\times n} \quad \mapsto \quad \mathcal{A}(\boldsymbol{M}) := \left\{\boldsymbol{a}_i^\top \boldsymbol{M} \boldsymbol{a}_i\right\}_{1\le i\le m}$$

and $\boldsymbol{A}$ the design matrix

$$\boldsymbol{A} := [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m]^\top.$$

2.1 Regularized gradient stage

For independent samples, the gradient of the real-valued Poisson log-likelihood obeys

$$\sum_{i=1}^m \nabla \ell_i(\boldsymbol{z}) = \sum_{i=1}^m 2 \underbrace{\frac{y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}}_{:=\nu_i} \boldsymbol{a}_i, \tag{21}$$

where ν_i represents the weight assigned to each $\boldsymbol{a}_i$. This forms the descent direction of WF updates.

Unfortunately, WF moving along the preceding direction might not come close to the truth unless $\boldsymbol{z}$ is already very close to $\boldsymbol{x}$. To see this, it is helpful to consider any fixed vector $\boldsymbol{z} \in \mathbb{R}^n$ independent of the design vectors. The typical size of $\min_{1\le i\le m} |\boldsymbol{a}_i^\top \boldsymbol{z}|$ is about on the order of $\frac{1}{m}\|\boldsymbol{z}\|$, introducing some unreasonably large weights ν_i, which can be as large as $m\|\boldsymbol{x}\|^2/\|\boldsymbol{z}\|$. Consequently, the iterative updates based on (1.2) often overshoot, and this arises starting from the very initial stage[6].

[6]For complex-valued data where $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) + j\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, WF converges empirically, as $\min_i |\boldsymbol{a}_i^* \boldsymbol{z}|$ is much larger than the real-valued case.

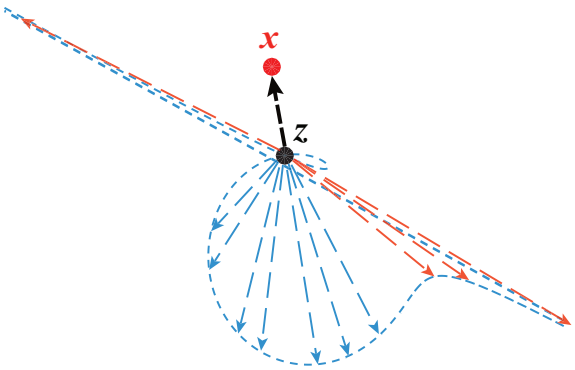


Figure 4: The locus of $-\frac{1}{2}\nabla \ell_i(\boldsymbol{z}) = \frac{|\boldsymbol{a}_i^\top \boldsymbol{z}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{x}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i$ when $\boldsymbol{a}_i$ ranges over all unit vectors, where $\boldsymbol{x} = (2.7, 8)$ and $\boldsymbol{z} = (3, 6)$. For each direction $\boldsymbol{a}_i$, $-\frac{1}{2}\nabla \ell_i(\boldsymbol{z})$ is aligned with $\boldsymbol{a}_i$, and its length represents the weight assigned to this component. In particular, the red arrows depict a few directions that behave like outliers, whereas the blue arrows depict several directions whose resulting gradients take typical sizes.

Fig. 4 illustrates this phenomenon by showing the locus of $-\nabla \ell_i(\boldsymbol{z})$ when $\boldsymbol{a}_i$ has unit norm and ranges over all possible directions. Examination of the figure seems to suggest that most of the gradient components $\nabla \ell_i(\boldsymbol{z})$ are more or less pointing towards the truth $\boldsymbol{x}$ and forming reasonable search directions. But there exist a few outlier components that are excessively large, which lead to unstable search directions. Notably, an underlying premise for a nonconvex procedure to succeed is to ensure all iterates reside within a *basin of attraction*, that is, a neighborhood surrounding $\boldsymbol{x}$ within which $\boldsymbol{x}$ is the unique stationary point of the objective. When a gradient is not well-controlled, the iterative procedure might overshoot and end up leaving this basin of attraction. This intuition is corroborated by numerical experiments under *real-valued* data. As illustrated in Fig. 5, the solutions returned by the WF (designed for a real-valued Poisson log-likelihood and $m = 8n$) are very far from the ground truth.

Hence, to remedy the aforementioned stability issue, it would be natural to separate the small fraction of abnormal gradient components by regularizing the weights ν_i, possibly via data-dependent trimming rules. This gives rise to the update rule of TWF:

$$\boldsymbol{z}^{(t+1)} \quad = \quad \boldsymbol{z}^{(t)} + \frac{\mu_t}{m} \nabla \ell_{\mathrm{tr}}(\boldsymbol{z}^{(t)}), \quad \forall t \in \mathbb{N}, \tag{22}$$

where $\nabla \ell_{\mathrm{tr}}(\cdot)$ denotes the regularized gradient given by[7]

$$\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}) := \sum_{i=1}^{m} 2 \frac{y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i(\boldsymbol{z}) \cap \mathcal{E}_2^i(\boldsymbol{z})}. \tag{23}$$

for some trimming criteria specified by $\mathcal{E}_1^i(\cdot)$ and $\mathcal{E}_2^i(\cdot)$. In our algorithm, we take $\mathcal{E}_1^i(\boldsymbol{z})$ and $\mathcal{E}_2^i(\boldsymbol{z})$ to be two collections of events given by

$$\mathcal{E}_1^i(\boldsymbol{z}) \quad := \quad \left\{ \alpha_z^{\mathrm{lb}} \le \frac{|\boldsymbol{a}_i^\top \boldsymbol{z}|}{\|\boldsymbol{z}\|} \le \alpha_z^{\mathrm{ub}} \right\}, \tag{24}$$

$$\mathcal{E}_2^i(\boldsymbol{z}) \quad := \quad \left\{ |y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2| \le \frac{\alpha_h}{m} \left\| \boldsymbol{y} - \mathcal{A}\left(\boldsymbol{z}\boldsymbol{z}^\top\right) \right\|_1 \frac{|\boldsymbol{a}_i^\top \boldsymbol{z}|}{\|\boldsymbol{z}\|} \right\}, \tag{25}$$

where α_z^{lb}, α_z^{ub}, α_z are predetermined thresholds. To keep notation light, we shall use $\mathcal{E}_1^i$ and $\mathcal{E}_2^i$ rather than $\mathcal{E}_1^i(\boldsymbol{z})$ and $\mathcal{E}_2^i(\boldsymbol{z})$ whenever it is clear from context.

[7] In the complex-valued case, the trimming rule is enforced upon the Wirtinger derivative, which reads $\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}) := \sum_{i=1}^{m} 2 \frac{y_i - |\boldsymbol{z}^* \boldsymbol{a}_i|^2}{\boldsymbol{z}^* \boldsymbol{a}_i} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i(\boldsymbol{z}) \cap \mathcal{E}_2^i(\boldsymbol{z})}$.

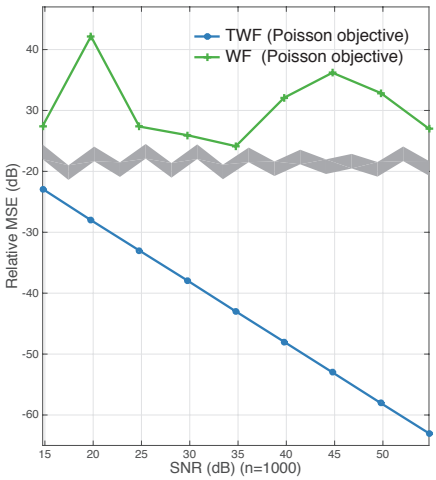


Figure 5: Relative MSE vs. SNR in dB. The curves are shown for WF and TWF, both employing the Poisson log-likelihood. Here, $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $n = 1000$, $m = 8n$, and each point is averaged over 100 Monte Carlo trials.

We emphasize that the above trimming procedure simply throws away those components whose weights ν_i's fall outside some confidence range, so as to remove the influence of outlier components. To achieve this, we regularize both the numerator and denominator of ν_i by enforcing separate trimming rules. Recognize that for any fixed $\boldsymbol{z}$, the denominator obeys

$$\mathbb{E}\left[\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|\right] = \sqrt{2/\pi}\|\boldsymbol{z}\|,$$

leading up to the rule (24). Regarding the numerator, by the law of large numbers one would expect

$$\mathbb{E}\left[\left|y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right|\right] \approx \frac{1}{m}\left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1,$$

and hence it is natural to regularize the numerator by ensuring

$$\left|y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \lesssim \frac{1}{m}\left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1.$$

As a remark, we include an extra term $\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|/\|\boldsymbol{z}\|$ in (25) to sharpen the theory, but all our results continue to hold (up to some modification of constants) if we drop this term in (25). Detailed procedures are summarized in Algorithm 1 [8].

The proposed paradigm could be counter-intuitive at first glance, since one might expect the larger terms to be better aligned with the desired search direction. The issue, however, is that the large terms are extremely volatile and could have too high of a leverage on the descent directions. In contrast, TWF discards these high-leverage data, which slightly increases the bias but remarkably reduces the variance of the descent direction. We expect such gradient regularization and variance reduction schemes to be beneficial for solving a broad family of nonconvex problems.

2.2 Truncated spectral initialization

In order for the gradient stage to converge rapidly, we need to seed it with a suitable initialization. One natural alternative is the spectral method adopted in [1, 21], which amounts to computing the leading

[8] Careful readers might note that we include some extra factor $\frac{\sqrt{n}}{\|\boldsymbol{a}_i\|}$ (which is approximately 1 in the Gaussian model) in Algorithm 1. This occurs since we present Algorithm 1 in a more general fashion that applies beyond the model $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, but all results / proofs continue to hold in the presence of this extra term.

Algorithm 1 Truncated Wirtinger Flow.

Input: Measurements $\{y_i \mid 1 \le i \le m\}$ and sampling vectors $\{\boldsymbol{a}_i \mid 1 \le i \le m\}$; trimming thresholds α_z^{lb}, α_z^{ub}, α_h, and α_y (see default values in Table 1).

Initialize $\boldsymbol{z}^{(0)}$ to be $\sqrt{\frac{mn}{\sum_{i=1}^m \|\boldsymbol{a}_i\|^2}}\lambda_0 \tilde{\boldsymbol{z}}$, where $\lambda_0 = \sqrt{\frac{1}{m}\sum_{i=1}^m y_i}$ and $\tilde{\boldsymbol{z}}$ is the leading eigenvector of

$$\boldsymbol{Y} = \frac{1}{m}\sum_{i=1}^m y_i \boldsymbol{a}_i \boldsymbol{a}_i^* \mathbf{1}_{\{|y_i| \le \alpha_y^2 \lambda_0^2\}}. \tag{26}$$

Loop: for $t = 0 : T$ **do**

$$\boldsymbol{z}^{(t+1)} \quad = \quad \boldsymbol{z}^{(t)} + \frac{2\mu_t}{m}\sum_{i=1}^m \frac{y_i - \left|\boldsymbol{a}_i^* \boldsymbol{z}^{(t)}\right|^2}{\boldsymbol{z}^{(t)*}\boldsymbol{a}_i} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i}, \tag{27}$$

where

$$\mathcal{E}_1^i := \left\{\alpha_z^{\mathrm{lb}} \le \frac{\sqrt{n}}{\|\boldsymbol{a}_i\|}\frac{|\boldsymbol{a}_i^* \boldsymbol{z}^{(t)}|}{\|\boldsymbol{z}^{(t)}\|} \le \alpha_z^{\mathrm{ub}}\right\}, \qquad \mathcal{E}_2^i := \left\{|y_i - |\boldsymbol{a}_i^* \boldsymbol{z}^{(t)}|^2| \le \alpha_h K_t \frac{\sqrt{n}}{\|\boldsymbol{a}_i\|}\frac{|\boldsymbol{a}_i^* \boldsymbol{z}^{(t)}|}{\|\boldsymbol{z}^{(t)}\|}\right\}, \tag{28}$$

$$\text{and} \quad K_t := \frac{1}{m}\sum_{l=1}^m \left|y_l - |\boldsymbol{a}_l^* \boldsymbol{z}^{(t)}|^2\right|.$$

Output $\boldsymbol{z}_T$.

eigenvector of $\widetilde{\boldsymbol{Y}} := \frac{1}{m}\sum_{i=1}^m y_i \boldsymbol{a}_i \boldsymbol{a}_i^\top$. This arises from the observation that when $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\|\boldsymbol{x}\| = 1$,

$$\mathbb{E}[\widetilde{\boldsymbol{Y}}] = \boldsymbol{I} + 2\boldsymbol{x}\boldsymbol{x}^\top,$$

whose leading eigenvector is exactly $\boldsymbol{x}$ with an eigenvalue of 3.

Unfortunately, this spectral technique converges to a good initial point only when $m \gtrsim n \log n$, due to the fact that $(\boldsymbol{a}_i^\top \boldsymbol{x})^2 \boldsymbol{a}_i \boldsymbol{a}_i^\top$ is heavy-tailed, a random quantity which does not have a moment generating function. To be more precise, consider the noiseless case $y_i = |\boldsymbol{a}_i^\top \boldsymbol{x}|^2$ and recall that $\max_i y_i \approx 2\log m$. Letting $k = \arg\max_i y_i$, one can calculate

$$\left(\frac{\boldsymbol{a}_k}{\|\boldsymbol{a}_k\|}\right)^\top \widetilde{\boldsymbol{Y}} \frac{\boldsymbol{a}_k}{\|\boldsymbol{a}_k\|} \ge \left(\frac{\boldsymbol{a}_k}{\|\boldsymbol{a}_k\|}\right)^\top \left(\frac{1}{m}\boldsymbol{a}_k \boldsymbol{a}_k^\top\right)\left(\boldsymbol{a}_k^\top \boldsymbol{x}\right)^2 \left(\frac{\boldsymbol{a}_k}{\|\boldsymbol{a}_k\|}\right) \approx \frac{2n\log m}{m},$$

which is much larger than $\boldsymbol{x}^\top \widetilde{\boldsymbol{Y}} \boldsymbol{x} = 3$ unless m/n is very large. This tells us that in the regime where $m \asymp n$, there exists some unit vector $\boldsymbol{a}_k/\|\boldsymbol{a}_k\|$ that is closer to the leading eigenvector of $\widetilde{\boldsymbol{Y}}$ than $\boldsymbol{x}$. This phenomenon happens because the summands of $\widetilde{\boldsymbol{Y}}$ have huge tails so that even one large term could end up dominating the empirical sum, thus preventing the spectral method from returning a meaningful initial guess.

To address this issue, we propose a more robust version of spectral method, which discards those observations y_i that are several times larger than the mean during spectral initialization. Specifically, the initial estimate is obtained by computing the leading eigenvector $\tilde{\boldsymbol{z}}$ of the truncated sum

$$\boldsymbol{Y} := \frac{1}{m}\sum_{i=1}^m y_i \boldsymbol{a}_i \boldsymbol{a}_i^\top \mathbf{1}_{\left\{|y_i| \le \alpha_y^2\left(\frac{1}{m}\sum_{l=1}^m y_l\right)\right\}} \tag{29}$$

for some predetermined threshold α_y, and then rescaling $\tilde{\boldsymbol{z}}$ so as to have roughly the same norm as $\boldsymbol{x}$ (which is estimated to be $\frac{1}{m}\sum_{l=1}^m y_l$); see Algorithm 1 for the detailed procedure.

Notably, the aforementioned drawback of the spectral method is not merely a theoretical concern but rather a substantial practical issue. We have seen this in Fig. 2 (main quad example) showing the enormous advantage of truncated spectral initialization. This is also further illustrated in Fig. 6, which compares the empirical efficiency of both methods with $\alpha_y = 3$ set to be the truncation threshold. For both Gaussian designs and CDP models, the empirical loss incurred by the original spectral method increases as n grows,

Table 1: Range of algorithmic parameters

(a) **When a fixed step size $\mu_t \equiv \mu$ is employed:** $(\alpha_z^{\mathrm{lb}}, \alpha_z^{\mathrm{ub}}, \alpha_h, \alpha_y)$ obeys

$$\begin{cases} \zeta_1 \;:= \max\left\{\mathbb{E}\left[\xi^2 \mathbf{1}_{\left\{|\xi| \le \sqrt{1.01}\alpha_z^{\mathrm{lb}} \text{ or } |\xi| \ge \sqrt{0.99}\alpha_z^{\mathrm{ub}}\right\}}\right], \mathbb{P}\left(|\xi| \le \sqrt{1.01}\alpha_z^{\mathrm{lb}} \text{ or } |\xi| \ge \sqrt{0.99}\alpha_z^{\mathrm{ub}}\right)\right\} \\ \zeta_2 \;:= \mathbb{E}\left[\xi^2 \mathbf{1}_{\{|\xi| > 0.473\alpha_h\}}\right], \\ 2(\zeta_1 + \zeta_2) + \sqrt{8/(9\pi)}\alpha_h^{-1} < 1.99, \\ \alpha_y \ge 3, \end{cases} \tag{30}$$

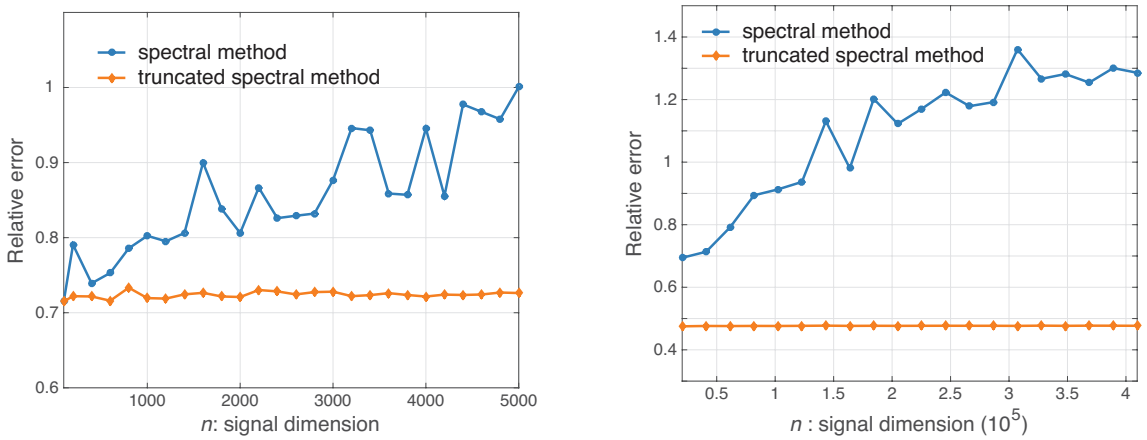
where $\xi \sim \mathcal{N}(0,1)$. By default, $\alpha_z^{\mathrm{lb}} = 0.3$, $\alpha_z^{\mathrm{ub}} = \alpha_h = 5$, and $\alpha_y = 3$.

(b) **When μ_t is chosen by a backtracking line search:** $(\alpha_z^{\mathrm{lb}}, \alpha_z^{\mathrm{ub}}, \alpha_h, \alpha_y, \alpha_p)$ obeys

$$0 < \alpha_z^{\mathrm{lb}} \le 0.1, \quad \alpha_z^{\mathrm{ub}} \ge 5, \quad \alpha_h \ge 6, \quad \alpha_y \ge 3, \quad \text{and} \quad \alpha_p \ge 5. \tag{31}$$

By default, $\alpha_z^{\mathrm{lb}} = 0.1$, $\alpha_z^{\mathrm{ub}} = 5$, $\alpha_h = 6$, $\alpha_y = 3$, and $\alpha_p = 5$.

which is in stark constrast to the truncated spectral method that achieves almost identical accuracy over the same range of n.



(a) (b)

Figure 6: The empirical relative error for both the spectral and the truncated spectral methods. The results are averaged over 50 Monte Carlo runs, and are shown for: (a) 1-D Gaussian measurement where $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $m = 6n$; (b) 2-D CDP model (9) where the diagonal entries of $\boldsymbol{D}^{(l)}$ are uniformly drawn from $\{1, -1, j, -j\}$, $n = n_1 \times n_2$ with $n_1 = 300$ and n_2 ranging from 64 to 1280, and $m = 12n$.

2.3 Choice of algorithmic parameters

One implementation detail to specify is the step size μ_t at each iteration t. There are two alternatives that work well in both theory and practice:

1. **Fixed step size**. Take $\mu_t \equiv \mu$ $(\forall t \in \mathbb{N})$ for some constant $\mu > 0$. As long as μ is not too large, our main results state that this strategy always works—although the convergence rate depends on μ. Under appropriate conditions, our theorems hold for any constant $0 < \mu < 0.28$.

2. **Backtracking line search with truncated objective**. This strategy performs a line search along the descent direction

$$\boldsymbol{p}_t := \frac{1}{m}\nabla \ell_{\rm tr}(\boldsymbol{z}_t)$$

and determines an appropriate step size that guarantees a sufficient improvement. In contrast to the conventional search strategy that determines the sufficient progress with respect to the true objective function, we propose to evaluate instead a regularized version of the objective function. Specifically, put

$$\widehat{\ell}(\boldsymbol{z}) := \sum_{i\in\widehat{\mathcal{T}}(\boldsymbol{z})} \left\{ y_i \log(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2) - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2 \right\}, \tag{32}$$

where

$$\widehat{\mathcal{T}}(\boldsymbol{z}) := \left\{ i \mid \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right| \geq \alpha_z^{\rm lb}\|\boldsymbol{z}\| \text{ and } \left|\boldsymbol{a}_i^\top \boldsymbol{p}\right| \leq \alpha_p\|\boldsymbol{p}\| \right\}.$$

Then the backtracking line search proceeds as

(a) Start with $\tau = 1$;

(b) Repeat $\tau \leftarrow \beta\tau$ until

$$\frac{1}{m}\widehat{\ell}\big(\boldsymbol{z}^{(t)} + \tau\boldsymbol{p}^{(t)}\big) \geq \frac{1}{m}\widehat{\ell}\big(\boldsymbol{z}^{(t)}\big) + \frac{1}{2}\tau\big\|\boldsymbol{p}^{(t)}\big\|^2, \tag{33}$$

where $\beta \in (0,1)$ is some pre-determined constant;

(c) Set $\mu_t = \tau$.

By definition (32), evaluating $\widehat{\ell}(\boldsymbol{z}^{(t)} + \tau\boldsymbol{p}^{(t)})$ mainly consists in calculating the matrix-vector product $\boldsymbol{A}(\boldsymbol{z}^{(t)} + \tau\boldsymbol{p}^{(t)})$. In total, we are going to evaluate $\widehat{\ell}(\boldsymbol{z}^{(t)} + \tau\boldsymbol{p}^{(t)})$ for $O\big(\log(1/\beta)\big)$ different τ's, and hence the total cost amounts to computing $\boldsymbol{A}\boldsymbol{z}^{(t)}$, $\boldsymbol{A}\boldsymbol{p}^{(t)}$ as well as $O(m\log(1/\beta))$ additional flops. Note that the matrix-vector products $\boldsymbol{A}\boldsymbol{z}^{(t)}$ and $\boldsymbol{A}\boldsymbol{p}^{(t)}$ need to be computed even when one adopts a pre-determined step size. Hence, the extra cost incurred by a backtracking line search, which is $O(m\log(1/\beta))$ flops, is negligible compared to that of computing the gradient even once.

Another set of important algorithmic parameters to determine is the trimming thresholds α_h, $\alpha_z^{\rm lb}$, $\alpha_z^{\rm ub}$, α_y, and α_p (for a backtracking line search only). The present paper isolates the set of $(\alpha_h, \alpha_z^{\rm lb}, \alpha_z^{\rm ub}, \alpha_y)$ obeying (30) as given in Table 1 when a fixed step size is employed. More concretely, this range subsumes as special cases all parameters obeying the following constraints:

$$0 < \alpha_z^{\rm lb} \leq 0.5, \quad \alpha_z^{\rm ub} \geq 5, \quad \alpha_h \geq 5, \quad \text{and} \quad \alpha_y \geq 3. \tag{34}$$

When a backtracking line search is adopted, an extra parameter α_p is needed, which we take to be $\alpha_p \geq 5$. In all theory presented herein, we assume that the parameters fall within the range singled out in Table 1.

3 Why TWF works?

Before proceeding, it is best to develop an intuitive understanding of the TWF iterations. We start with a notation representing the (unrecoverable) global phase [1] for real-valued data

$$\phi\left(\boldsymbol{z}\right) := \begin{cases} 0, & \text{if } \|\boldsymbol{z}-\boldsymbol{x}\| \leq \|\boldsymbol{z}+\boldsymbol{x}\|, \\ \pi, & \text{else.} \end{cases} \tag{35}$$

It is self-evident that

$$(-\boldsymbol{z}) + \frac{\mu}{m}\nabla_{\rm tr}\ell\big(-\boldsymbol{z}\big) = -\left\{\boldsymbol{z} + \frac{\mu}{m}\nabla_{\rm tr}\ell(\boldsymbol{z})\right\},$$

and hence (cf. Definition (8))

$$\operatorname{dist}\left((-\boldsymbol{z}) + \frac{\mu}{m}\nabla_{\rm tr}\ell(-\boldsymbol{z}), \boldsymbol{x}\right) = \operatorname{dist}\left(\boldsymbol{z} + \frac{\mu}{m}\nabla_{\rm tr}\ell\left(\boldsymbol{z}\right), \boldsymbol{x}\right)$$

despite the global phase uncertainty. For simplicity of presentation, we shall drop the phase term by letting $\boldsymbol{z}$ be $e^{-j\phi(\boldsymbol{z})}\boldsymbol{z}$ and setting $\boldsymbol{h} = \boldsymbol{z} - \boldsymbol{x}$, whenever it is clear from context.

The first object to consider is the descent direction. To this end, we find it convenient to work with a fixed $\boldsymbol{z}$ independent of the design vectors $\boldsymbol{a}_i$, which is of course heuristic but helpful in developing some intuition. Rewrite

$$\begin{aligned}\nabla \ell_i(\boldsymbol{z}) &= 2\frac{(\boldsymbol{a}_i^\top \boldsymbol{x})^2 - (\boldsymbol{a}_i^\top \boldsymbol{z})^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}\boldsymbol{a}_i \overset{\text{(i)}}{=} -2\frac{(\boldsymbol{a}_i^\top \boldsymbol{h})(2\boldsymbol{a}_i^\top \boldsymbol{z} - \boldsymbol{a}_i^\top \boldsymbol{h})}{\boldsymbol{a}_i^\top \boldsymbol{z}}\boldsymbol{a}_i \\ &= -4(\boldsymbol{a}_i^\top \boldsymbol{h})\boldsymbol{a}_i + \underbrace{2\frac{(\boldsymbol{a}_i^\top \boldsymbol{h})^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}\boldsymbol{a}_i}_{:=\boldsymbol{r}_i}, \end{aligned} \tag{36}$$

where (i) follows from the identity $a^2 - b^2 = (a+b)(a-b)$. The first component of (36), which on average gives $-4\boldsymbol{h}$, makes a good search direction when averaged over all the observations $i = 1, \ldots, m$. The issue is that the other term $\boldsymbol{r}_i$—which is in general non-integrable—could be devastating. The reason is that $\boldsymbol{a}_i^\top \boldsymbol{z}$ could be arbitrarily small, thus resulting in an unbounded $\boldsymbol{r}_i$. As a consequence, a non-negligible portion of the $\boldsymbol{r}_i$'s may exert a very strong influence on the descent direction in an undesired manner.

Such an issue can be prevented if one can detect and separate those gradient components bearing abnormal $\boldsymbol{r}_i$'s. Since we cannot observe the individual components of the decomposition (36), we cannot reject indices with large values of $\boldsymbol{r}_i$ directly. Instead, we examine each gradient component as a whole and discard it if its size is not absolutely controlled. Fortunately, such a strategy is sufficient to ensure that most of the contribution from the regularized gradient comes from the first component of (36), namely, $-4(\boldsymbol{a}_i^\top \boldsymbol{h})\boldsymbol{a}_i$. As will be made precise in Proposition 2 and Lemma 7, the regularized gradient obeys

$$-\Big\langle \frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}), \boldsymbol{h}\Big\rangle \;\geq\; (4-\epsilon)\|\boldsymbol{h}\|^2 - O\bigg(\frac{\|\boldsymbol{h}\|^3}{\|\boldsymbol{z}\|}\bigg) \tag{37}$$

$$\text{and} \quad \Big\|\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\| \lesssim \|\boldsymbol{h}\|. \tag{38}$$

Here, one has $(4-\epsilon)\|\boldsymbol{h}\|^2$ in (37) instead of $4\|\boldsymbol{h}\|^2$ to account for the bias introduced by adaptive trimming, where ϵ is small as long as we only throw away a small fraction of data. Looking at (37) and (38) we see that the search direction is sufficiently aligned with the deviation $-\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{z}$ of the current iterate; i.e. they form a reasonably good angle that is bounded away from 90°. Consequently, $\boldsymbol{z}$ is expected to be dragged towards $\boldsymbol{x}$ provided that the step size is appropriately chosen.

The observations (37) and (38) are reminiscent of a (local) regularity condition given in [1], which is a fundamental criterion that dictates rapid convergence of iterative procedures (including WF and other gradient descent schemes). When specialized to TWF, we say that $-\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\cdot)$ satisfies the *regularity condition*, denoted by $\mathsf{RC}(\mu, \lambda, \epsilon)$, if

$$\Big\langle \boldsymbol{h}, -\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\rangle \;\geq\; \frac{\mu}{2}\Big\|\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\|^2 + \frac{\lambda}{2}\|\boldsymbol{h}\|^2 \tag{39}$$

holds for all $\boldsymbol{z}$ obeying $\|\boldsymbol{z} - \boldsymbol{x}\| \leq \epsilon\|\boldsymbol{x}\|$, where $0 < \epsilon < 1$ is some constant. Such an ϵ-ball around $\boldsymbol{x}$ forms a basin of attraction. Formally, under $\mathsf{RC}(\mu, \lambda, \epsilon)$, a little algebra gives

$$\begin{aligned}\mathrm{dist}^2\Big(\boldsymbol{z} + \frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}), \boldsymbol{x}\Big) &\leq \Big\|\boldsymbol{z} + \frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}) - \boldsymbol{x}\Big\|^2 \\ &= \|\boldsymbol{h}\|^2 + \Big\|\frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\|^2 + 2\mu\Big\langle \boldsymbol{h}, \frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\rangle \\ &\leq \|\boldsymbol{h}\|^2 + \Big\|\frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\|^2 - \mu^2\Big\|\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\Big\|^2 - \mu\lambda\|\boldsymbol{h}\|^2 \\ &= (1-\mu\lambda)\,\mathrm{dist}^2(\boldsymbol{z}, \boldsymbol{x}) \end{aligned} \tag{40}$$

for any $\boldsymbol{z}$ with $\|\boldsymbol{z} - \boldsymbol{x}\| \leq \epsilon$. In words, the TWF update rule is locally contractive around the planted solution, provided that $\mathsf{RC}(\mu, \lambda, \epsilon)$ holds for some nonzero μ and λ. Apparently, Conditions (37) and (38)

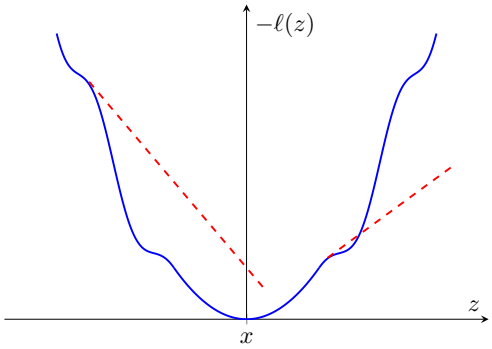


Figure 7: A function $-\ell(z)$ satisfying RC: $-\ell(z) = z^2$ for any $z \in [-6, 6]$, and $-\ell(z) = z^2 + 1.5|z|\,(\cos(|z| - 6) - 1)$ otherwise.

already imply the validity of RC for some constants $\mu, \lambda \asymp 1$ when $\|\boldsymbol{h}\|/\|\boldsymbol{z}\|$ is reasonably small, which in turn allows us to take a constant step size μ and enables a constant contraction rate $1 - \mu\lambda$.

Finally, caution must be exercised when connecting RC with strong convexity, since the former does not necessarily guarantee the latter within the basin of attraction. As an illustration, Fig. 7 plots the graph of a non-convex function obeying RC. The distinction stems from the fact that RC is stated only for those pairs $\boldsymbol{z}$ and $\boldsymbol{h} = \boldsymbol{z} - \boldsymbol{x}$ with $\boldsymbol{x}$ being a fixed component, rather than simultaneously accommodating all possible $\boldsymbol{z}$ and $\boldsymbol{h} = \boldsymbol{z} - \tilde{\boldsymbol{z}}$ with $\tilde{\boldsymbol{z}}$ being an arbitrary vector. In contrast, RC says that the only stationary point of the truncated objective in a neighborhood of $\boldsymbol{x}$ is $\boldsymbol{x}$, which often suffices for a gradient-descent type scheme to succeed.

4 Numerical experiments

In this section, we report additional numerical results to verify the practical applicability of TWF. In all numerical experiments conducted in the current paper, we set

$$\alpha_z^{\mathrm{lb}} = 0.3, \quad \alpha_z^{\mathrm{ub}} = 5, \quad \alpha_h = 5, \quad \text{and} \quad \alpha_y = 3. \tag{41}$$

This is a concrete combination of parameters satisfying our condition (30). Unless otherwise noted, we employ 50 power iterations for initialization, adopt a fixed step size $\mu_t \equiv 0.2$ when updating TWF iterates, and set the maximum number of iterations to be $T = 1000$ for the iterative refinement stage.

The first series of experiments concerns exact recovery from noise-free data. Set $n = 1000$ and generate a real-valued signal $\boldsymbol{x}$ at random. Then for m varying between $2n$ and $6n$, generate m design vectors $\boldsymbol{a}_i$ independently drawn from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. An experiment is claimed to succeed if the returned estimate $\hat{\boldsymbol{x}}$ satisfies $\mathrm{dist}\,(\hat{\boldsymbol{x}}, \boldsymbol{x}) \,/\, \|\boldsymbol{x}\| \le 10^{-5}$. Fig. 8 illustrates the empirical success rate of TWF (over 100 Monte Carlo trials for each m) revealing that exact recovery is practially guaranteed from fewer than 1000 iterations when the number of quadratic constraints is about 5 times the ambient dimension.

To see how special the real-valued Gaussian designs are to our theoretical finding, we perform experiments on two other types of measurement models. In the first, TWF is applied to complex-valued data by generating $\boldsymbol{a}_i \sim \mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}\right) + j\mathcal{N}\left(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}\right)$. The other is the model of coded diffraction patterns described in (9). Fig. 9 depicts the average success rate for both types of measurements over 100 Monte Carlo trials, indicating that $m > 4.5n$ and $m \ge 6n$ are often sufficient under complex-valued Gaussian and CDP models, respectively.

For the sake of comparison, we also report the empirical performance of WF in all the above settings, where the step size is set to be the default choice of [1], that is, $\mu_t = \min\{1 - e^{-t/330}, 0.2\}$. As can be seen, the empirical success rates of TWF outperform WF when $T = 1000$ under Gaussian models, suggesting that TWF either converges faster or exhibits better phase transition behavior.

Another series of experiments has been carried out to demonstrate the stability of TWF when the number m of quadratic equations varies. We consider the case where $n = 1000$, and vary the SNR (cf. (10)) from 15 dB to 55dB. The design vectors are real-valued independent Gaussian $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, while the measurements

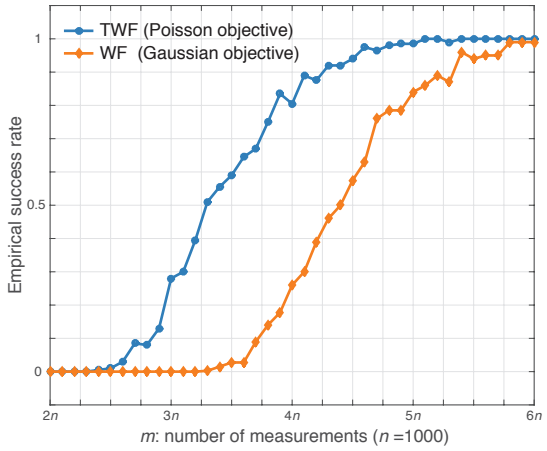


Figure 8: Empirical success rate under real-valued Gaussian sampling $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.

y_i are generated according to the Poisson noise model (4). Fig. 10 shows the relative mean square error—in the dB scale—as a function of SNR, when averaged over 100 independent runs. For all choices of m, the numerical experiments demonstrate that the relative MSE scales inversely proportional to SNR, which matches our stability guarantees in Theorem 2 (since we observe that on the dB scale, the slope is about -1 as predicted by the theory (19)).

5 Exact recovery from noiseless data

This section proves the theoretical guarantees of TWF in the absence of noise (i.e. Theorem 1). We separate the noiseless case mainly out of pedagogical reasons, as most of the steps carry over to the noisy case with slight modification.

The analysis for the truncated spectral method relies on the celebrated Davis-Kahan $\sin\Theta$ theorem [56], which we defer to Appendix C. In short, for any fixed $\delta > 0$ and $\boldsymbol{x} \in \mathbb{R}^n$, the initial point $\boldsymbol{z}^{(0)}$ returned by the truncated spectral method obeys

$$\operatorname{dist}(\boldsymbol{z}^{(0)}, \boldsymbol{x}) \le \delta \|\boldsymbol{x}\|$$

with high probability, provided that m/n exceeds some numerical constant. With this in place, it suffices to demonstrate that the TWF update rule is locally contractive, as stated in the following proposition.

Proposition 1 (Local error contraction). *Consider the noiseless case (1). Under the condition (30), there exist some universal constants $0 < \rho_0 < 1$ and $c_0, c_1, c_2 > 0$ such that with probability exceeding $1 - c_1 \exp(-c_2 m)$,*

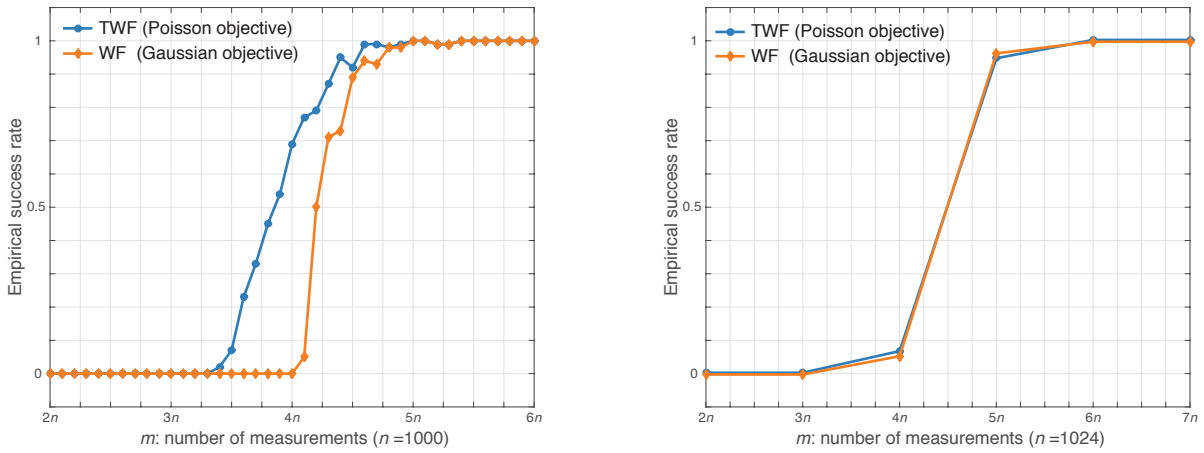
$$\operatorname{dist}^2\left(\boldsymbol{z} + \frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}), \boldsymbol{x}\right) \quad \le \quad (1-\rho_0)\operatorname{dist}^2(\boldsymbol{z}, \boldsymbol{x}) \tag{42}$$

holds simultaneously for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ obeying

$$\frac{\operatorname{dist}(\boldsymbol{z}, \boldsymbol{x})}{\|\boldsymbol{z}\|} \quad \le \quad \min\left\{\frac{1}{11}, \frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h}, \frac{\alpha_z^{\mathrm{lb}}}{6}, \frac{5.7\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}} + \alpha_z^{\mathrm{lb}}}\right\}, \tag{43}$$

provided that $m \ge c_0 n$ and that μ is some constant obeying

$$0 < \mu \le \mu_0 := \frac{0.994 - \zeta_1 - \zeta_2 - \sqrt{2/(9\pi)}\alpha_h^{-1}}{2\left(1.02 + 0.665/\alpha_h\right)}.$$



(a) (b)

Figure 9: Empirical success rate for exact recovery using TWF. The results are shown for (a) complex-valued Gaussian sampling $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n) + j\mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_n)$, and (b) CDP with masks uniformly drawn from $\{1, -1, j, -j\}$.

Proposition 1 reveals the monotonicity of the estimation error: once entering a neighborhood around $\boldsymbol{x}$ of a reasonably small size, the iterative updates will remain within this neighborhood all the time and be attracted towards $\boldsymbol{x}$ at a geometric rate.

As shown in Section 3, under the hypothesis $\mathsf{RC}(\mu, \lambda, \epsilon)$ one can conclude

$$\mathrm{dist}^2\left(\boldsymbol{z} + \frac{\mu}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}), \boldsymbol{x}\right) \le (1 - \mu\lambda)\mathrm{dist}^2(\boldsymbol{z}, \boldsymbol{x}), \quad \forall (\boldsymbol{z}, \boldsymbol{x}) \text{ with } \mathrm{dist}(\boldsymbol{z}, \boldsymbol{x}) \le \epsilon. \tag{44}$$

Thus, everything now boils down to showing $\mathsf{RC}(\mu, \lambda, \epsilon)$ for some constants $\mu, \lambda, \epsilon > 0$. This occupies the rest of this section.

5.1 Preliminary facts about $\{\mathcal{E}_1^i\}$ and $\{\mathcal{E}_2^i\}$

Before proceeding, we gather a few properties of the events $\mathcal{E}_1^i$ and $\mathcal{E}_2^i$, which will prove crucial in establishing $\mathsf{RC}(\mu, \lambda, \epsilon)$. To begin with, recall that the truncation level given in $\mathcal{E}_2^i$ depends on $\frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1$. Instead of working with this random variable directly, we use deterministic quantities that are more amenable to analysis. Specifically, we claim that $\frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1$ offers a uniform and orderwise tight estimate on $\|\boldsymbol{h}\|\|\boldsymbol{z}\|$, which can be seen from the following two facts.

Lemma 1. *Fix $\zeta \in (0, 1)$. If $m > c_0 n\zeta^{-2}\log\frac{1}{\zeta}$, then with probability at least $1 - C\exp(-c_1\zeta^2 m)$,*

$$0.9\,(1-\zeta)\,\|\boldsymbol{M}\|_{\mathrm{F}} \le \frac{1}{m}\,\|\mathcal{A}(\boldsymbol{M})\|_1 \le (1+\zeta)\,\sqrt{2}\,\|\boldsymbol{M}\|_{\mathrm{F}} \tag{45}$$

holds for all symmetric rank-2 matrices $\boldsymbol{M} \in \mathbb{R}^{n\times n}$. Here, $c_0, c_1, C > 0$ are some universal constants.

Proof. Since [6, Lemma 3.1] already establishes the upper bound, it suffices to prove the lower tail bound. Consider all symmetric rank-2 matrices $\boldsymbol{M}$ with eigenvalues 1 and $-t$ for some $-1 \le t \le 1$. When $t \in [0, 1]$, it has been shown in the proof of [6, Lemma 3.2] that with high probability,

$$\frac{1}{m}\,\|\mathcal{A}(\boldsymbol{M})\|_1 \ge (1-\zeta)\,f(t)\,, \tag{46}$$

for all such rank-2 matrices $\boldsymbol{M}$, where $f(t) := \frac{2}{\pi}\left\{2\sqrt{t} + (1-t)\left(\pi/2 - 2\arctan(\sqrt{t})\right)\right\}$. The lower bound in this case can then be justified by recognizing that $f(t)/\sqrt{1+t^2} \ge 0.9$ for all $t \in [0, 1]$, as illustrated in Fig. 11. The case where $t \in [-1, 0]$ is an immediate consequence from [6, Lemma 3.1]. □

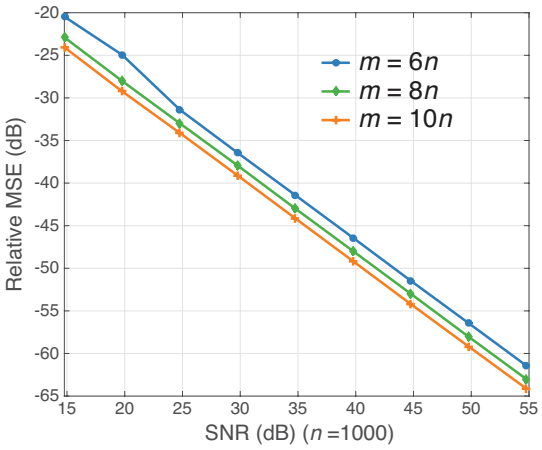


Figure 10: Relative MSE vs. SNR when the y_i's follow the Poisson model.

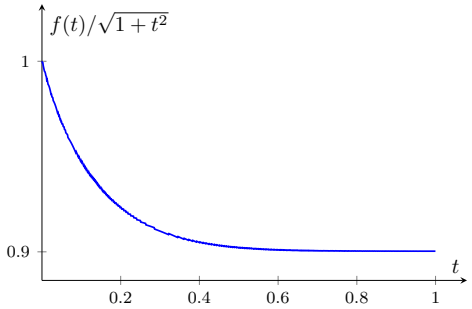


Figure 11: $\frac{f(t)}{\sqrt{1+t^2}}$ as a function of t.

Lemma 2. *Consider any $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ obeying $\|\boldsymbol{z}-\boldsymbol{x}\| \le \delta \|\boldsymbol{z}\|$ for some $\delta < \frac{1}{2}$. Then one has*

$$\sqrt{2-4\delta}\,\|\boldsymbol{z}-\boldsymbol{x}\|\,\|\boldsymbol{z}\| \le \left\|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right\|_{\mathrm{F}} \le (2+\delta)\,\|\boldsymbol{z}-\boldsymbol{x}\|\,\|\boldsymbol{z}\|\,. \tag{47}$$

Proof. Take $\boldsymbol{h} = \boldsymbol{z} - \boldsymbol{x}$ and write

$$\begin{aligned}\left\|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right\|_{\mathrm{F}}^2 &= \left\| -\boldsymbol{h}\boldsymbol{z}^\top - \boldsymbol{z}\boldsymbol{h}^\top + \boldsymbol{h}\boldsymbol{h}^\top\right\|_{\mathrm{F}}^2 \\ &= \left\|\boldsymbol{h}\boldsymbol{z}^\top + \boldsymbol{z}\boldsymbol{h}^\top\right\|_{\mathrm{F}}^2 + \|\boldsymbol{h}\|^4 - 2\langle \boldsymbol{h}\boldsymbol{z}^\top + \boldsymbol{z}\boldsymbol{h}^\top, \boldsymbol{h}\boldsymbol{h}^\top\rangle \\ &= 2\,\|\boldsymbol{z}\|^2\,\|\boldsymbol{h}\|^2 + 2|\boldsymbol{h}^\top\boldsymbol{z}|^2 + \|\boldsymbol{h}\|^4 - 2\|\boldsymbol{h}\|^2(\boldsymbol{h}^\top\boldsymbol{z} + \boldsymbol{z}^\top\boldsymbol{h}).\end{aligned}$$

When $\|\boldsymbol{h}\| < \frac{1}{2}\|\boldsymbol{z}\|$, the Cauchy-Schwarz inequality gives

$$2\,\|\boldsymbol{z}\|^2\,\|\boldsymbol{h}\|^2 - 4\,\|\boldsymbol{z}\|\,\|\boldsymbol{h}\|^3 \le \left\|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right\|_{\mathrm{F}}^2 \le 4\,\|\boldsymbol{z}\|^2\,\|\boldsymbol{h}\|^2 + 4\,\|\boldsymbol{h}\|^3\,\|\boldsymbol{z}\| + \|\boldsymbol{h}\|^4\,, \tag{48}$$

$$\Rightarrow \quad \sqrt{(2\,\|\boldsymbol{z}\| - 4\,\|\boldsymbol{h}\|)\,\|\boldsymbol{z}\|}\cdot\|\boldsymbol{h}\| \le \left\|\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right\|_{\mathrm{F}} \le (2\,\|\boldsymbol{z}\| + \|\boldsymbol{h}\|)\cdot\|\boldsymbol{h}\| \tag{49}$$

as claimed. □

Taken together the above two facts demonstrate that with probability $1 - \exp(-\Omega(m))$,

$$1.15\,\|\boldsymbol{z}-\boldsymbol{x}\|\,\|\boldsymbol{z}\| \le \frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1 \le 3\,\|\boldsymbol{z}-\boldsymbol{x}\|\,\|\boldsymbol{z}\| \tag{50}$$

holds simultaneously for all $\boldsymbol{z}$ and $\boldsymbol{x}$ satisfying $\|\boldsymbol{h}\| \le \frac{1}{11}\|\boldsymbol{z}\|$. Conditional on (50), the inclusion

$$\mathcal{E}_3^i \;\subseteq\; \mathcal{E}_2^i \;\subseteq\; \mathcal{E}_4^i \tag{51}$$

holds with respect to the following events

$$\mathcal{E}_3^i : \;=\; \left\{\left||\boldsymbol{a}_i^\top \boldsymbol{x}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \le 1.15\alpha_h \|\boldsymbol{h}\| \cdot \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|\right\}, \tag{52}$$

$$\mathcal{E}_4^i : \;=\; \left\{\left||\boldsymbol{a}_i^\top \boldsymbol{x}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \le 3\alpha_h \|\boldsymbol{h}\| \cdot \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|\right\}. \tag{53}$$

The point of introducing these new events is that the $\mathcal{E}_3^i$'s (resp. $\mathcal{E}_4^i$'s) are statistically independent for any fixed $\boldsymbol{x}$ and $\boldsymbol{z}$ and are, therefore, easier to work with.

Note that each $\mathcal{E}_3^i$ (resp. $\mathcal{E}_4^i$) is specified by a quadratic inequality. A closer inspection reveals that in order to satisfy these quadratic inequalities, the quantity $\boldsymbol{a}_i^\top \boldsymbol{h}$ must fall within two intervals centered around 0 and $2\boldsymbol{a}_i^\top \boldsymbol{z}$, respectively. One can thus facilitate analysis by decoupling each quadratic inequality of interest into two simple linear inequalities, as stated in the following lemma.

Lemma 3. *For any $\gamma > 0$, define*

$$\mathcal{D}_\gamma^i \;:=\; \left\{\left||\boldsymbol{a}_i^\top \boldsymbol{x}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \le \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|\right\}, \tag{54}$$

$$\mathcal{D}_\gamma^{i,1} \;:=\; \left\{\frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|}{\|\boldsymbol{h}\|} \le \gamma\right\}, \tag{55}$$

$$\text{and} \quad \mathcal{D}_\gamma^{i,2} \;:=\; \left\{\left|\frac{\boldsymbol{a}_i^\top \boldsymbol{h}}{\|\boldsymbol{h}\|} - \frac{2\boldsymbol{a}_i^\top \boldsymbol{z}}{\|\boldsymbol{h}\|}\right| \le \gamma\right\}. \tag{56}$$

Thus, $\mathcal{D}_\gamma^{i,1}$ and $\mathcal{D}_\gamma^{i,2}$ represent the two intervals on $\boldsymbol{a}_i^\top \boldsymbol{h}$ centered around 0 and $2\boldsymbol{a}_i^\top \boldsymbol{z}$. If $\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \frac{\alpha_z^{\mathrm{lb}}}{\gamma}$, then the following inclusion holds

$$\left(\mathcal{D}^{i,1}_{\frac{\gamma}{1+\sqrt{2}}} \cap \mathcal{E}_1^i\right) \cup \left(\mathcal{D}^{i,2}_{\frac{\gamma}{1+\sqrt{2}}} \cap \mathcal{E}_1^i\right) \subseteq \;\; \mathcal{D}_\gamma^i \cap \mathcal{E}_1^i \;\; \subseteq \;\; \left(\mathcal{D}_\gamma^{i,1} \cap \mathcal{E}_1^i\right) \cup \left(\mathcal{D}_\gamma^{i,2} \cap \mathcal{E}_1^i\right). \tag{57}$$

5.2 Proof of the regularity condition

By definition, one step towards proving the regularity condition (39) is to control the norm of the regularized gradient. In fact, a crude argument already reveals that $\|\frac{1}{m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\| \lesssim \|\boldsymbol{h}\|$. To see this, introduce $\boldsymbol{v} = [v_i]_{1\le i\le m}$ with $v_i := 2\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i}$. It comes from the trimming rule $\mathcal{E}_1^i$ as well as the inclusion property (51) that

$$\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right| \gtrsim \|\boldsymbol{z}\| \quad \text{and} \quad \left|y_i - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2\right| \lesssim \frac{1}{m}\|\mathcal{A}(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top)\|_1 \asymp \|\boldsymbol{h}\|\,\|\boldsymbol{z}\|,$$

implying $|v_i| \lesssim \|\boldsymbol{h}\|$ and hence $\|\boldsymbol{v}\| \lesssim \sqrt{m}\|\boldsymbol{h}\|$. The Marchenko–Pastur law gives $\|\boldsymbol{A}\| \lesssim \sqrt{m}$, whence

$$\frac{1}{m}\|\nabla \ell_{\mathrm{tr}}(\boldsymbol{z})\| = \frac{1}{m}\|\boldsymbol{A}^\top \boldsymbol{v}\| \le \frac{1}{m}\|\boldsymbol{A}\| \cdot \|\boldsymbol{v}\| \lesssim \|\boldsymbol{h}\|. \tag{58}$$

A more refined estimate will be provided in Lemma 7.

The above argument essentially tells us that to establish RC, it suffices to verify a uniform lower bound of the form

$$-\left\langle \boldsymbol{h},\; \frac{1}{m}\nabla \ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\rangle \;\gtrsim\; \|\boldsymbol{h}\|^2, \tag{59}$$

as formally derived in the following proposition.

Proposition 2. *Consider the noise-free measurements $y_i = |\boldsymbol{a}_i^\top \boldsymbol{x}|^2$ and any fixed constant $\epsilon > 0$. Under the condition (30), if $m > c_1 n$, then with probability exceeding $1 - C\exp\left(-c_0 m\right)$,*

$$-\left\langle \boldsymbol{h}, \frac{1}{m}\nabla \ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\rangle \;\ge\; 2\left\{1.99 - 2\left(\zeta_1 + \zeta_2\right) - \sqrt{8/(9\pi)}\alpha_h^{-1} - \epsilon\right\}\|\boldsymbol{h}\|^2 \tag{60}$$

holds uniformly over all $\boldsymbol{x},\, \boldsymbol{z} \in \mathbb{R}^n$ *obeying*

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \;\le\; \min\left\{\frac{1}{11},\; \frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h},\; \frac{\alpha_z^{\mathrm{lb}}}{6},\; \frac{5.7\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}}+\alpha_z^{\mathrm{lb}}}\right\}. \tag{61}$$

Here, $c_0, c_1, C > 0$ *are some universal constants, and* ζ_1 *and* ζ_2 *are defined in (30).*

The basic starting point is the observation that $(\boldsymbol{a}_i^\top \boldsymbol{z}) - (\boldsymbol{a}_i^\top \boldsymbol{x})^2 = (\boldsymbol{a}_i^\top \boldsymbol{h})(2\boldsymbol{a}_i^\top \boldsymbol{z} - \boldsymbol{a}_i^\top \boldsymbol{h})$ and hence

$$\begin{aligned} -\frac{1}{2m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}) &= \frac{1}{m}\sum_{i=1}^m \frac{(\boldsymbol{a}_i^\top \boldsymbol{z})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x})^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} \\ &= \frac{1}{m}\sum_{i=1}^m 2(\boldsymbol{a}_i^\top \boldsymbol{h})\boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} - \frac{1}{m}\sum_{i=1}^m \frac{(\boldsymbol{a}_i^\top \boldsymbol{h})^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i}. \end{aligned} \tag{62}$$

One would expect the contribution of the second term of (62) (which is a second-order quantity) to be small as $\|\boldsymbol{h}\| / \|\boldsymbol{z}\|$ decreases.

To facilitate analysis, we rewrite (62) in terms of the more convenient events $\mathcal{D}_\gamma^{i,1}$ and $\mathcal{D}_\gamma^{i,2}$. Specifically, the inclusion property (51) together with Lemma 3 reveals that

$$\mathcal{D}_{\gamma_3}^{i,1} \cap \mathcal{E}_1^i \;\subseteq\; \mathcal{E}_3^i \cap \mathcal{E}_1^i \subseteq\; \mathcal{E}_2^i \cap \mathcal{E}_1^i \;\subseteq\; \mathcal{E}_4^i \cap \mathcal{E}_1^i \;\subseteq\; \left(\mathcal{D}_{\gamma_4}^{i,1} \cup \mathcal{D}_{\gamma_4}^{i,2}\right) \cap \mathcal{E}_1^i, \tag{63}$$

where the parameters γ_3, γ_4 are given by

$$\gamma_3 := 0.476\alpha_h, \quad \text{and} \quad \gamma_4 := 3\alpha_h. \tag{64}$$

This taken collectively with the identity (62) leads to a lower estimate

$$\begin{aligned} &-\left\langle \frac{1}{2m}\nabla \ell_{\mathrm{tr}}(\boldsymbol{z}), \boldsymbol{h}\right\rangle \ge \\ &\quad \frac{2}{m}\sum_{i=1}^m \left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_3}^{i,1}} - \frac{1}{m}\sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|} \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_4}^{i,1}} - \frac{1}{m}\sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|} \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_4}^{i,2}}, \end{aligned} \tag{65}$$

leaving us with three quantities in the right-hand side to deal with. We pause here to explain and compare the influences of these three terms.

To begin with, as long as the trimming step does not discard too many data, the first term should be close to $\frac{2}{m}\sum_i |\boldsymbol{a}_i^\top \boldsymbol{h}|^2$, which approximately gives $2\|\boldsymbol{h}\|^2$ from the law of large numbers. This term turns out to be dominant in the right-hand side of (65) as long as $\|\boldsymbol{h}\|/\|\boldsymbol{z}\|$ is reasonably small. To see this, please recognize that the second term in the right-hand side is $O(\|\boldsymbol{h}\|^3/\|\boldsymbol{z}\|)$, simply because both $\boldsymbol{a}_i^\top \boldsymbol{h}$ and $\boldsymbol{a}_i^\top \boldsymbol{z}$ are absolutely controlled on $\mathcal{D}_{\gamma_4}^{i,1} \cap \mathcal{E}_1^i$. However, $\mathcal{D}_{\gamma_4}^{i,2}$ does not share such a desired feature. By the very definition of $\mathcal{D}_{\gamma_4}^{i,2}$, each nonzero summand of the last term of (65) must obey $\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right| \approx 2\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|$ and, therefore, $\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|} \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_4}^{i,2}}$ is roughly of the order of $\|\boldsymbol{z}\|^2$; this could be much larger than our target level $\|\boldsymbol{h}\|^2$. Fortunately, $\mathcal{D}_{\gamma_4}^{i,2}$ is a rare event, thus precluding a noticable influence upon the descent direction. All of this is made rigorous in Lemma 4 (first term), Lemma 5 (second term) and Lemma 6 (third term) together with subsequent analysis.

Lemma 4. *Fix* $\gamma > 0$*, and let* $\mathcal{E}_1^i$ *and* $\mathcal{D}_\gamma^{i,1}$ *be defined in (24) and (55), respectively. Set*

$$\zeta_1 := 1 - \min\left\{\mathbb{E}\left[\xi^2 \mathbf{1}_{\left\{\sqrt{1.01}\alpha_z^{\mathrm{lb}} \le |\xi| \le \sqrt{0.99}\alpha_z^{\mathrm{ub}}\right\}}\right], \mathbb{E}\left[\mathbf{1}_{\left\{\sqrt{1.01}\alpha_z^{\mathrm{lb}} \le |\xi| \le \sqrt{0.99}\alpha_z^{\mathrm{ub}}\right\}}\right]\right\} \tag{66}$$

and

$$\zeta_2 := \mathbb{E}\left[\xi^2 \mathbf{1}_{\left\{|\xi| > \sqrt{0.99}\gamma\right\}}\right], \tag{67}$$

where $\xi \sim \mathcal{N}(0,1)$*. For any* $\epsilon > 0$*, if* $m > c_1 n \epsilon^{-2}\log\epsilon^{-1}$*, then with probability at least* $1 - C\exp(-c_0\epsilon^2 m)$*,*

$$\frac{1}{m}\sum_{i=1}^m \left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2 \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1}} \;\;\ge\;\; (1 - \zeta_1 - \zeta_2 - \epsilon)\|\boldsymbol{h}\|^2 \tag{68}$$

holds for all non-zero vectors $\boldsymbol{h}, \boldsymbol{z} \in \mathbb{R}^n$*. Here,* $c_0, c_1, C > 0$ *are some universal constants.*

We now move on to the second term in the right-hand side of (65). For any fixed $\gamma > 0$, the definition of $\mathcal{E}_1^i$ gives rise to an upper estimate

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{D}_\gamma^{i,1}} \le \frac{1}{\alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\|}\frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3\mathbf{1}_{\mathcal{D}_\gamma^{i,1}} \le \frac{(1+\epsilon)\sqrt{8/\pi}\left\|\boldsymbol{h}\right\|^3}{\alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\|}, \tag{69}$$

where $\sqrt{8/\pi}\left\|\boldsymbol{h}\right\|^3$ is exactly the untruncated moment $\mathbb{E}[|\boldsymbol{a}_i^\top \boldsymbol{h}|^3]$. The second inequality is a consequence of the lemma below, which arises by observing that the summands $|\boldsymbol{a}_i^\top \boldsymbol{h}|^3\mathbf{1}_{\mathcal{D}_\gamma^{i,1}}$ are independent sub-Gaussian random variables.

Lemma 5. *For any constant $\gamma > 0$, if $m/n \ge c_0 \cdot \epsilon^{-2}\log\epsilon^{-1}$, then*

$$\frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3\mathbf{1}_{\mathcal{D}_\gamma^{i,1}} \le (1+\epsilon)\sqrt{8/\pi}\left\|\boldsymbol{h}\right\|^3, \quad \forall \boldsymbol{h}\in\mathbb{R}^n \tag{70}$$

with probability at least $1 - C\exp(-c_1\epsilon^2 m)$ for some universal constants $c_0, c_1, C > 0$.

It remains to control the last term of (65). As mentioned above, the influence of this term is small since the set of $\boldsymbol{a}_i$'s satisfying $\mathcal{D}_\gamma^{i,2}$ accounts for a small fraction of measurements. Put formally, the number of equations satisfying $\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right| \ge \gamma\left\|\boldsymbol{h}\right\|$ decays rapidly for large γ (at least at a quadratic rate), as stated below.

Lemma 6. *For any $0 < \epsilon < 1$, there exist some universal constants $c_0, c_1, C > 0$ such that*

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}|\ge\,\gamma\|\boldsymbol{h}\|\right\}} \le \frac{1}{0.49\gamma}\exp\left(-0.485\gamma^2\right) + \frac{\epsilon}{\gamma^2}, \quad \forall \boldsymbol{h}\in\mathbb{R}^n\backslash\{\boldsymbol{0}\} \text{ and } \gamma \ge 2 \tag{71}$$

with probability at least $1 - C\exp\left(-c_0\epsilon^2 m\right)$. This holds with the proviso $m/n \ge c_1\cdot\epsilon^{-2}\log\epsilon^{-1}$.

To connect this lemma with the last term of (65), we recognize that when $\gamma \le \frac{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}{\|\boldsymbol{h}\|}$, one has

$$\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{D}_\gamma^{i,2}} \quad\le\quad \mathbf{1}_{\left\{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|\ge\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|\right\}}. \tag{72}$$

The constraint $\left|\frac{\boldsymbol{a}_i^\top \boldsymbol{h}}{\|\boldsymbol{h}\|} - \frac{2\boldsymbol{a}_i^\top \boldsymbol{z}}{\|\boldsymbol{h}\|}\right| \le \gamma$ of $\mathcal{D}_\gamma^{i,2}$ necessarily requires

$$\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|}{\|\boldsymbol{h}\|} \ge \frac{2\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\|\boldsymbol{h}\|} - \gamma \ge \frac{2\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}{\|\boldsymbol{h}\|} - \gamma \ge \frac{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}{\|\boldsymbol{h}\|}, \tag{73}$$

where the last inequality comes from our assumption on γ. With Lemma 6 in place, (72) immediately gives

$$\begin{aligned}\sum_{i=1}^{m}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{D}_\gamma^{i,2}} &\le \frac{\|\boldsymbol{h}\|}{0.49\alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\|}\exp\left(-0.485\left(\frac{\alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\|}{\|\boldsymbol{h}\|}\right)^2\right) + \frac{\epsilon\left\|\boldsymbol{h}\right\|^2}{\left(\alpha_z^{\mathrm{lb}}\right)^2\left\|\boldsymbol{z}\right\|^2} \\ &\le \frac{1}{9800}\left(\frac{\|\boldsymbol{h}\|}{\alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\|}\right)^4 + \frac{\epsilon}{\left(\alpha_z^{\mathrm{lb}}\right)^2}\left(\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)^2\end{aligned} \tag{74}$$

as long as $\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \frac{\alpha_z^{\mathrm{lb}}}{6}$, where the last inequality uses the majorization $\frac{1}{20000x^4} \ge \frac{1}{x}\exp\left(-0.485x^2\right)$ holding for any $x \ge 6$.

In addition, on $\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,2}$, the amplitude of each summand can be bounded in such a way that

$$\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|} \le \frac{\left|2\boldsymbol{a}_i^\top \boldsymbol{z}\right| + \gamma\left\|\boldsymbol{h}\right\|}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\left(2\alpha_z^{\mathrm{ub}}\left\|\boldsymbol{z}\right\| + \gamma\left\|\boldsymbol{h}\right\|\right)^2 \tag{75}$$

$$\le \left(2 + \frac{\gamma}{\alpha_z^{\mathrm{lb}}}\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)\left(2\alpha_z^{\mathrm{ub}} + \gamma\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)^2\left\|\boldsymbol{z}\right\|^2, \tag{76}$$

where both inequalities are immediate consequences from the definitions of $\mathcal{D}_\gamma^{i,2}$ and $\mathcal{E}_1^i$ (see (56) and (24)). Taking this together with the cardinality bound (74) and picking ϵ appropriately, we get

$$\frac{1}{m}\sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{D}_\gamma^{i,2}} \le \left\{ \underbrace{\frac{\left(2+\frac{\gamma}{\alpha_z^{\mathrm{lb}}}\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)\left(2\alpha_z^{\mathrm{ub}}+\gamma\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)^2}{9800\left(\alpha_z^{\mathrm{lb}}\right)^4}}_{\vartheta_1}\frac{\|\boldsymbol{h}\|^2}{\|\boldsymbol{z}\|^2}+\epsilon \right\}\|\boldsymbol{h}\|^2. \tag{77}$$

Furthermore, under the condition that

$$\gamma \le \alpha_z^{\mathrm{lb}}\frac{\|\boldsymbol{z}\|}{\|\boldsymbol{h}\|} \quad \text{and} \quad \frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \frac{\sqrt{98}\left(\alpha_z^{\mathrm{lb}}\right)^2}{\sqrt{3}\left(2\alpha_z^{\mathrm{ub}}+\alpha_z^{\mathrm{lb}}\right)},$$

one can simplify (77) by observing that $\vartheta_1 \le \frac{1}{100}$, which results in

$$\frac{1}{m}\sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{D}_\gamma^{i,2}} \quad \le \quad \left(\frac{1}{100}+\epsilon\right)\|\boldsymbol{h}\|^2. \tag{78}$$

Putting all preceding results in this subsection together reveals that with probability exceeding $1-\exp\left(-\Omega\left(m\right)\right)$,

$$\begin{aligned} -\left\langle \boldsymbol{h}, \frac{1}{2m}\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\rangle &\ge \left\{1.99-2\left(\zeta_1+\zeta_2\right)-\sqrt{8/\pi}\frac{\|\boldsymbol{h}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}-3\epsilon\right\}\|\boldsymbol{h}\|^2 \\ &\ge \left\{1.99-2\left(\zeta_1+\zeta_2\right)-\sqrt{8/\pi}(3\alpha_h)^{-1}-3\epsilon\right\}\|\boldsymbol{h}\|^2 \end{aligned} \tag{79}$$

holds simultaneously over all $\boldsymbol{x}$ and $\boldsymbol{z}$ satisfying

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \min\left\{\frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h},\ \frac{\alpha_z^{\mathrm{lb}}}{6},\ \frac{\sqrt{98/3}\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}}+\alpha_z^{\mathrm{lb}}},\frac{1}{11}\right\} \tag{80}$$

as claimed in Proposition 2.

To conclude this section, we provide a tighter estimate about the norm of the regularized gradient.

Lemma 7. *Fix $\delta>0$, and assume that $y_i=(\boldsymbol{a}_i^\top\boldsymbol{x})^2$. Suppose that $m\ge c_0 n$ for some large constant $c_0>0$. There exist some universal constants $c,C>0$ such that with probability at least $1-C\exp\left(-cm\right)$,*

$$\frac{1}{m}\left\|\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\| \ \le\ (1+\delta)\cdot 4\sqrt{1.02+0.665/\alpha_h}\,\|\boldsymbol{h}\| \tag{81}$$

holds simultaneously for all $\boldsymbol{x},\ \boldsymbol{z}\in\mathbb{R}^n$ satisfying $\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\le\min\left\{\frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h},\frac{\alpha_z^{\mathrm{lb}}}{6},\frac{\sqrt{98/3}\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}}+\alpha_z^{\mathrm{lb}}},\frac{1}{11}\right\}$.

Lemma 7 complements the preceding arguments by allowing us to identify a concrete plausible range for the step size. Specifically, putting Lemma 7 and Proposition 2 together suggests that

$$-\left\langle \boldsymbol{h}, \frac{1}{m}\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\rangle \ \ge\ \frac{2\left\{1.99-2\left(\zeta_1+\zeta_2\right)-\sqrt{8/(9\pi)}\alpha_h^{-1}-\epsilon\right\}}{\left(1+\delta\right)^2\cdot 16\left(1.02+0.665/\alpha_h\right)}\left\|\frac{1}{m}\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right)\right\|^2. \tag{82}$$

Taking ϵ and δ to be sufficiently small we arrive at a feasible range (cf. Definition (39))

$$\mu \le \frac{0.994-\zeta_1-\zeta_2-\sqrt{2/(9\pi)}\alpha_h^{-1}}{2\left(1.02+0.665/\alpha_h\right)} := \mu_0. \tag{83}$$

This establishes Proposition 1 and in turn Theorem 1 when μ_t is taken to be a fixed constant.

To justify the contraction under backtracking line search, it suffices to prove that the resulting step size falls within this range (83), which we defer to Appendix D.

6 Stability

This section goes in the direction of establishing stability guarantees of TWF. We concentrate on the iterative gradient stage, and defer the analysis for the initialization stage to Appendix C.

Before continuing, we collect two bounds that we shall use several times. The first is the observation that

$$\begin{aligned}\frac{1}{m}\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{z}\boldsymbol{z}^\top)\|_1 &\le \frac{1}{m}\|\mathcal{A}(\boldsymbol{x}\boldsymbol{x}^\top-\boldsymbol{z}\boldsymbol{z}^\top)\|_1+\frac{1}{m}\|\boldsymbol{\eta}\|_1\\ &\lesssim \|\boldsymbol{h}\|\|\boldsymbol{z}\|+\frac{1}{m}\|\boldsymbol{\eta}\|_1 \lesssim \|\boldsymbol{h}\|\|\boldsymbol{z}\|+\frac{1}{\sqrt{m}}\|\boldsymbol{\eta}\|,\end{aligned} \tag{84}$$

where the last inequality follows from Cauchy-Schwarz. Setting

$$v_i := 2\frac{y_i-|\boldsymbol{a}_i^\top\boldsymbol{z}|^2}{\boldsymbol{a}_i^\top\boldsymbol{z}}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}$$

as usual, this inequality together with the trimming rules $\mathcal{E}_1^i$ and $\mathcal{E}_2^i$ gives

$$\begin{aligned}&|v_i|\lesssim \|\boldsymbol{h}\|+\tfrac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|}\\ \Longrightarrow\quad \big\|\tfrac{1}{m}\nabla\ell_{\mathrm{tr}}(\boldsymbol{z})\big\| &= \tfrac{1}{m}\|\boldsymbol{A}^\top\boldsymbol{v}\| \le \Big\|\tfrac{1}{\sqrt{m}}\boldsymbol{A}\Big\|\tfrac{1}{\sqrt{m}}\|\boldsymbol{v}\| \overset{\text{(i)}}{\lesssim} \tfrac{1}{\sqrt{m}}\|\boldsymbol{v}\| \lesssim \|\boldsymbol{h}\|+\tfrac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|},\end{aligned} \tag{85}$$

where (i) arises from [40, Corollary 5.35].

As discussed in Section 3, the estimation error is contractive if $-\frac{1}{m}\nabla\ell_{\mathrm{tr}}(\boldsymbol{z})$ satisfies the regularity condition. With (85) in place, RC reduces to

$$-\frac{1}{m}\langle\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}),\boldsymbol{h}\rangle \gtrsim \|\boldsymbol{h}\|^2. \tag{86}$$

Unfortunately, (86) does not hold for all $\boldsymbol{z}$ within the neighborhood of $\boldsymbol{x}$ due to the existence of noise. Instead we establish the following:

- The condition (86) holds for all $\boldsymbol{h}$ obeying

$$c_3\frac{\|\boldsymbol{\eta}\|/\sqrt{m}}{\|\boldsymbol{z}\|}\le\|\boldsymbol{h}\|\le c_4\|\boldsymbol{x}\| \tag{87}$$

  for some constants $c_3,c_4>0$ (we shall call it *Regime 1*); this will be proved later. In this regime, the reasoning in Section 3 gives

$$\mathrm{dist}\Big(\boldsymbol{z}+\frac{\mu}{m}\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}),\ \boldsymbol{x}\Big)\le(1-\rho)\mathrm{dist}(\boldsymbol{z},\boldsymbol{x}) \tag{88}$$

  for some appropriate constants $\mu,\rho>0$ and, hence, error contraction occurs as in the noiseless setting.

- However, once the iterate enters *Regime 2* where

$$\|\boldsymbol{h}\|\le\frac{c_3\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|}, \tag{89}$$

  the estimation error might no longer be contractive. Fortunately, in this regime each move by $\frac{\mu}{m}\nabla\ell_{\mathrm{tr}}(\boldsymbol{z})$ is of size at most $O(\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|})$, compare (85). As a result, at each iteration the estimation error cannot increase by more than a numerical constant times $\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|}$ before possibly jumping out (of this regime). Therefore,

$$\mathrm{dist}\big(\boldsymbol{z}+\frac{\mu}{m}\nabla\ell_{\mathrm{tr}}(\boldsymbol{z}),\ \boldsymbol{x}\big)\le c_5\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{x}\|} \tag{90}$$

  for some constant $c_5>0$. Moreover, as long as $\|\boldsymbol{\eta}\|_\infty/\|\boldsymbol{x}\|^2$ is sufficiently small, one can guarantee that $c_5\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{x}\|}\le c_5\frac{\|\boldsymbol{\eta}\|_\infty}{\|\boldsymbol{x}\|}\le c_4\|\boldsymbol{x}\|$. In other words, if the iterate jumps out of Regime 2, it will still fall within Regime 1.

To summarize, suppose the initial guess $\boldsymbol{z}^{(0)}$ obeys $\mathrm{dist}(\boldsymbol{z}^{(0)},\boldsymbol{x}) \le c_4\|\boldsymbol{x}\|$. Then the estimation error will shrink at a geometric rate $1-\rho$ before it enters Regime 2. Afterwards, $\boldsymbol{z}^{(t)}$ will either stay within Regime 2 or jump back and forth between Regimes 1 and 2. Because of the bounds (90) and (88), the estimation errors will never exceed the order of $\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{x}\|}$ from then on. Putting these together establishes (16), namely, the first part of the theorem.

Below we justify the condition (86) for Regime 1, for which we start by gathering additional properties of the trimming rules. By Cauchy-Schwarz, $\frac{1}{m}\|\boldsymbol{\eta}\|_1 \le \frac{1}{\sqrt{m}}\|\boldsymbol{\eta}\| \le \frac{1}{c_3}\|\boldsymbol{h}\|\|\boldsymbol{z}\|$. When c_3 is sufficiently large, applying Lemmas 1 and 2 gives

$$\begin{array}{l} \frac{1}{m}\sum_{l=1}^m \left| y_l - \left|\boldsymbol{a}_l^\top \boldsymbol{z}\right|^2 \right| \le \frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1 + \frac{1}{m}\|\boldsymbol{\eta}\|_1 \le 2.98\|\boldsymbol{h}\|\|\boldsymbol{z}\|; \\ \frac{1}{m}\sum_{l=1}^m \left| y_l - \left|\boldsymbol{a}_l^\top \boldsymbol{z}\right|^2 \right| \ge \frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{x}\boldsymbol{x}^\top - \boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1 - \frac{1}{m}\|\boldsymbol{\eta}\|_1 \ge 1.151\|\boldsymbol{h}\|\|\boldsymbol{z}\|. \end{array} \tag{91}$$

From now on, we shall denote $\tilde{\mathcal{E}}_2^i := \left\{ \left||\boldsymbol{a}_i^\top \boldsymbol{x}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \le \frac{\alpha_h}{m}\left\|\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{z}\boldsymbol{z}^\top\right)\right\|_1 \frac{|\boldsymbol{a}_i^\top \boldsymbol{z}|}{\|\boldsymbol{z}\|} \right\}$ to differentiate from $\mathcal{E}_2^i$. For any small constant $\epsilon > 0$, we introduce the index set $\mathcal{G} := \{i : |\eta_i| \le C_\epsilon \|\boldsymbol{\eta}\|/\sqrt{m}\}$ that satisfies $|\mathcal{G}| = (1-\epsilon)m$. Note that C_ϵ must be bounded as n scales, since

$$\|\boldsymbol{\eta}\|^2 \;\ge\; \sum\nolimits_{i\notin\mathcal{G}} \eta_i^2 \;\ge\; (m - |\mathcal{G}|)\cdot C_\epsilon^2 \|\boldsymbol{\eta}\|^2/m \;\ge\; \epsilon C_\epsilon^2 \|\boldsymbol{\eta}\|^2 \quad\Rightarrow\quad C_\epsilon \le 1/\sqrt{\epsilon}. \tag{92}$$

We are now ready to analyze the regularized gradient, which we separate into several components as follows

$$\begin{aligned} \nabla_{\mathrm{tr}}\ell(\boldsymbol{z}) = \; & \underbrace{2\sum_{i\in\mathcal{G}} \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} + 2\sum_{i\notin\mathcal{G}} \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i\cap\tilde{\mathcal{E}}_2^i}}_{:=\nabla_{\mathrm{tr}}^{\mathrm{clean}}\ell(\boldsymbol{z})} \\ & + \underbrace{2\sum_{i\in\mathcal{G}} \frac{\eta_i}{\boldsymbol{a}_i^\top \boldsymbol{z}} \boldsymbol{a}_i \mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}}_{:=\nabla_{\mathrm{tr}}^{\mathrm{noise}}\ell(\boldsymbol{z})} + \underbrace{2\sum_{i\notin\mathcal{G}} \left(\frac{y_i - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} - \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}} \mathbf{1}_{\mathcal{E}_1^i\cap\tilde{\mathcal{E}}_2^i} \right) \boldsymbol{a}_i}_{:=\nabla_{\mathrm{tr}}^{\mathrm{extra}}\ell(\boldsymbol{z})}. \end{aligned} \tag{93}$$

- For each index $i \in \mathcal{G}$, the inclusion property (51) (i.e. $\mathcal{E}_3^i \subseteq \mathcal{E}_2^i \subseteq \mathcal{E}_4^i$) holds. To see this, observe that

$$\left|y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \in \left[\left||\boldsymbol{a}_i^\top \boldsymbol{x}|^2 - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right| \pm |\eta_i|\right].$$

  Since $|\eta_i| \le C_\epsilon\|\boldsymbol{\eta}\|/\sqrt{m} \ll \|\boldsymbol{h}\|\|\boldsymbol{z}\|$ when c_3 is sufficiently large, one can derive the inclusion (51) immediately from (91). As a result, all the proof arguments for Proposition 2 carry over to $\nabla_{\mathrm{tr}}^{\mathrm{clean}}\ell(\boldsymbol{z})$, suggesting that

$$-\left\langle \boldsymbol{h}, \frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{clean}}\ell(\boldsymbol{z}) \right\rangle \;\ge\; 2\left\{1.99 - 2(\zeta_1+\zeta_2) - \sqrt{8/(9\pi)}\alpha_h^{-1} - \epsilon\right\}\|\boldsymbol{h}\|^2. \tag{94}$$

- Next, letting $w_i = \frac{2\eta_i}{\boldsymbol{a}_i^\top \boldsymbol{z}} \mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}\mathbf{1}_{\{i\in\mathcal{G}\}}$, we see that for any constant $\delta > 0$, the noise component obeys

$$\begin{aligned} \left\|\frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{noise}}\ell(\boldsymbol{z})\right\| \;&=\; \left\|\frac{1}{m}\boldsymbol{A}^\top \boldsymbol{w}\right\| \;\le\; \left\|\frac{1}{\sqrt{m}}\boldsymbol{A}\right\| \left\|\frac{1}{\sqrt{m}}\boldsymbol{w}\right\| \\ &\overset{\text{(ii)}}{\le}\; \frac{1+\delta}{\sqrt{m}}\|\boldsymbol{w}\| \;\le\; (1+\delta)\frac{2\|\boldsymbol{\eta}\|/\sqrt{m}}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}, \end{aligned} \tag{95}$$

  provided that m/n is sufficiently large. Here, (ii) arises from [40, Corollary 5.35], and the last inequality is a consequence of the upper estimate

$$\|\boldsymbol{w}\|^2 \;\le 4\sum_{i=1}^m \frac{|\eta_i|^2}{(\boldsymbol{a}_i^\top \boldsymbol{z})^2}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} \;\le\; 4\sum_{i=1}^m \frac{|\eta_i|^2}{(\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|)^2} \;=\; \frac{4\|\boldsymbol{\eta}\|^2}{(\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|)^2}. \tag{96}$$

In turn, this immediately gives

$$\left|\left\langle \boldsymbol{h}, \frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{noise}}\ell\left(\boldsymbol{z}\right)\right\rangle\right| \;\le\; \|\boldsymbol{h}\|\left\|\frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{noise}}\ell\left(\boldsymbol{z}\right)\right\| \;\le\; \frac{2\left(1+\delta\right)}{\alpha_z^{\mathrm{lb}}}\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|}\|\boldsymbol{h}\|. \tag{97}$$

- We now turn to the last term $\nabla_{\mathrm{tr}}^{\mathrm{extra}}\ell\left(\boldsymbol{z}\right)$. According to the definition of $\mathcal{E}_2^i$ and $\tilde{\mathcal{E}}_2^i$ as well as the property (91), the weight $q_i := 2\Big(\frac{y_i-\left|\boldsymbol{a}_i^\top\boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top\boldsymbol{z}}\mathbf{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} - \frac{\left|\boldsymbol{a}_i^\top\boldsymbol{x}\right|^2-\left|\boldsymbol{a}_i^\top\boldsymbol{z}\right|^2}{\boldsymbol{a}_i^\top\boldsymbol{z}}\mathbf{1}_{\mathcal{E}_1^i\cap\tilde{\mathcal{E}}_2^i}\Big)\mathbf{1}_{\{i\notin\mathcal{G}\}}$ is bounded in magnitude by $6\|\boldsymbol{h}\|$. This gives

$$\|\boldsymbol{q}\| \le \sqrt{m-|\mathcal{G}|}\cdot 6\|\boldsymbol{h}\| \le 6\sqrt{\epsilon m}\|\boldsymbol{h}\|,$$

and hence

$$\left|\left\langle \frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{extra}}\ell\left(\boldsymbol{z}\right), \boldsymbol{h}\right\rangle\right| \le \|\boldsymbol{h}\|\cdot\Big\|\frac{1}{m}\nabla_{\mathrm{tr}}^{\mathrm{extra}}\ell\left(\boldsymbol{z}\right)\Big\| = \frac{1}{m}\|\boldsymbol{h}\|\cdot\left\|\boldsymbol{A}^\top\boldsymbol{q}\right\| \le 6\left(1+\delta\right)\sqrt{\epsilon}\|\boldsymbol{h}\|^2. \tag{98}$$

Taking the above bounds together yields

$$\begin{aligned}-\frac{1}{m}\left\langle\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right),\boldsymbol{h}\right\rangle \;\ge\; & 2\left\{1.99-2\left(\zeta_1+\zeta_2\right)-\sqrt{\frac{8}{9\pi}}\frac{1}{\alpha_h}-6(1+\delta)\sqrt{\epsilon}-\epsilon\right\}\|\boldsymbol{h}\|^2 \\ & -\frac{2\left(1+\delta\right)}{\alpha_z^{\mathrm{lb}}}\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\,\|\boldsymbol{z}\|}\|\boldsymbol{h}\|.\end{aligned}$$

Since $\|\boldsymbol{h}\| \ge c_3\frac{\|\boldsymbol{\eta}\|}{\sqrt{m}\|\boldsymbol{z}\|}$ for some large constant $c_3>0$, setting ϵ to be small one obtains

$$-\frac{1}{m}\left\langle\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right),\boldsymbol{h}\right\rangle \;\;\ge\;\; 2\left\{1.95-2\left(\zeta_1+\zeta_2\right)-\sqrt{8/(9\pi)}\alpha_h^{-1}\right\}\|\boldsymbol{h}\|^2 \tag{99}$$

for all $\boldsymbol{h}$ obeying

$$\frac{c_3\|\boldsymbol{\eta}\|/\sqrt{m}}{\|\boldsymbol{z}\|} \;\le\; \|\boldsymbol{h}\| \;\le\; \min\left\{\frac{1}{11}, \frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h}, \frac{\alpha_z^{\mathrm{lb}}}{6}, \frac{\sqrt{98/3}\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}}+\alpha_z^{\mathrm{lb}}}\right\}\|\boldsymbol{z}\|,$$

which finishes the proof of Theorem 2 for general $\boldsymbol{\eta}$.

Up until now, we have established the theorem for general $\boldsymbol{\eta}$, and it remains to specialize it to the Poisson model. Standard concentration results, which we omit, give

$$\frac{1}{m}\|\boldsymbol{\eta}\|^2 \approx \frac{1}{m}\sum_{i=1}^m\mathbb{E}\left[\eta_i^2\right] = \frac{1}{m}\sum_{i=1}^m\left(\boldsymbol{a}_i^\top\boldsymbol{x}\right)^2 \approx \|\boldsymbol{x}\|^2 \tag{100}$$

with high probability. Substitution into (16) completes the proof.

7 Minimax lower bound

The goal of this section is to establish the minimax lower bound given in Theorem 3. For notational simplicity, we denote by $\mathbb{P}\left(\boldsymbol{y}\mid\boldsymbol{w}\right)$ the likelihood of $y_i \overset{\text{ind.}}{\sim} \mathsf{Poisson}(|\boldsymbol{a}_i^\top\boldsymbol{w}|^2)$, $1\le i\le m$ conditional on $\{\boldsymbol{a}_i\}$. For any two probability measures P and Q, we denote by $\mathsf{KL}\left(P\|Q\right)$ the Kullback–Leibler (KL) divergence between them:

$$\mathsf{KL}\left(P\|Q\right) := \int\log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}P, \tag{101}$$

The basic idea is to adopt the general reduction scheme discussed in [41, Section 2.2], which amounts to finding a finite collection of hypotheses that are minimally separated. Below we gather one result useful for constructing and analyzing such hypotheses.

Lemma 8. *Suppose that $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$, n is sufficiently large, and $m = \kappa n$ for some sufficiently large constant $\kappa > 0$. Consider any $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$. On an event $\mathcal{B}$ of probability approaching one, there exists a collection $\mathcal{M}$ of $M = \exp(n/30)$ distinct vectors obeying the following properties:*

(i) $\boldsymbol{x} \in \mathcal{M}$;

(ii) for all $\boldsymbol{w}^{(l)}, \boldsymbol{w}^{(j)} \in \mathcal{M}$,

$$1/\sqrt{8} - (2n)^{-1/2} \le \left\|\boldsymbol{w}^{(l)} - \boldsymbol{w}^{(j)}\right\| \le 3/2 + n^{-1/2}; \tag{102}$$

(iii) for all $\boldsymbol{w} \in \mathcal{M}$,

$$\frac{|\boldsymbol{a}_i^\top (\boldsymbol{w} - \boldsymbol{x})|^2}{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2} \le \frac{\|\boldsymbol{w} - \boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}\{2 + 17\log^3 m\}, \quad 1 \le i \le m. \tag{103}$$

In words, Lemma 8 constructs a set $\mathcal{M}$ of exponentially many vectors/hypotheses scattered around $\boldsymbol{x}$ and yet well separated. From (ii) we see that each pair of hypotheses in $\mathcal{M}$ is separated by a distance roughly on the order of 1, and all hypotheses reside within a spherical ball centered at $\boldsymbol{x}$ of radius $3/2 + o(1)$. When $\|\boldsymbol{x}\| \ge \log^{1.5} m$, every hypothesis $\boldsymbol{w} \in \mathcal{M}$ satisfies $\|\boldsymbol{w}\| \approx \|\boldsymbol{x}\| \gg 1$. In addition, (iii) says that the quantities $|\boldsymbol{a}_i^\top (\boldsymbol{w} - \boldsymbol{x})|/|\boldsymbol{a}_i^\top \boldsymbol{x}|$ are all very well controlled (modulo some logarithmic factor). In particular, when $\|\boldsymbol{x}\| \ge \log^{1.5} m$, one must have

$$\frac{|\boldsymbol{a}_i^\top (\boldsymbol{w} - \boldsymbol{x})|^2}{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2} \lesssim \frac{\|\boldsymbol{w} - \boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}\log^3 m \lesssim \frac{1}{\log^3 m}\log^3 m \lesssim 1. \tag{104}$$

In the Poisson model, such a quantity turns out to be crucial in controlling the information divergence between two hypotheses, as demonstrated in the following lemma.

Lemma 9. *Fix a family of design vectors $\{\boldsymbol{a}_i\}$. Then for any $\boldsymbol{w}$ and $\boldsymbol{r} \in \mathbb{R}^n$,*

$$\mathsf{KL}\big(\mathbb{P}(\boldsymbol{y} \mid \boldsymbol{w} + \boldsymbol{r}) \parallel \mathbb{P}(\boldsymbol{y} \mid \boldsymbol{w})\big) \;\le\; \sum\nolimits_{i=1}^m |\boldsymbol{a}_i^\top \boldsymbol{r}|^2\left(8 + \frac{2|\boldsymbol{a}_i^\top \boldsymbol{r}|^2}{|\boldsymbol{a}_i^\top \boldsymbol{w}|^2}\right). \tag{105}$$

Lemma 9 and (104) taken collectively suggest that on the event $\mathcal{B} \cap \mathcal{C}$ ($\mathcal{B}$ is in Lemma 8 and $\mathcal{C} := \{\|\boldsymbol{A}\| \le \sqrt{2m}\}$), the conditional KL divergence (we condition on the $\boldsymbol{a}_i$'s) obeys

$$\mathsf{KL}\big(\mathbb{P}(\boldsymbol{y} \mid \boldsymbol{w}) \parallel \mathbb{P}(\boldsymbol{y} \mid \boldsymbol{x})\big) \le c_3 \sum\nolimits_{i=1}^m \left|\boldsymbol{a}_i^\top (\boldsymbol{w} - \boldsymbol{x})\right|^2 \le 2c_3 m \|\boldsymbol{w} - \boldsymbol{x}\|^2, \quad \forall \boldsymbol{w} \in \mathcal{M}; \tag{106}$$

here, the inequality holds for some constant $c_3 > 0$ provided that $\|\boldsymbol{x}\| \ge \log^{1.5} m$, and the last inequality is a result of $\mathcal{C}$ (which occurs with high probability). We now use hypotheses as in Lemma 8 but rescaled in such a way that

$$\|\boldsymbol{w} - \boldsymbol{x}\| \asymp \delta, \quad \text{and} \quad \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\| \asymp \delta, \quad \forall \boldsymbol{w}, \tilde{\boldsymbol{w}} \in \mathcal{M} \text{ with } \boldsymbol{w} \ne \tilde{\boldsymbol{w}}. \tag{107}$$

for some $0 < \delta < 1$. This is achieved via the substitution $\boldsymbol{w} \longleftarrow \boldsymbol{x} + \delta(\boldsymbol{w} - \boldsymbol{x})$; with a slight abuse of notation, $\mathcal{M}$ denotes the new set.

The hardness of a minimax estimation problem is known to be dictated by information divergence inequalities such as (106). Indeed, suppose that

$$\frac{1}{M-1}\sum\nolimits_{\boldsymbol{w} \in \mathcal{M}\backslash\{\boldsymbol{x}\}} \mathsf{KL}\big(\mathbb{P}(\boldsymbol{y} \mid \boldsymbol{w}) \parallel \mathbb{P}(\boldsymbol{y} \mid \boldsymbol{x})\big) \le \frac{1}{10}\log(M-1) \tag{108}$$

holds, then the Fano-type minimax lower bound [41, Theorem 2.7] asserts that

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x} \in \mathcal{M}} \mathbb{E}\left[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| \;\middle|\; \{\boldsymbol{a}_i\}\right] \gtrsim \min_{\boldsymbol{w}, \tilde{\boldsymbol{w}} \in \mathcal{M}, \boldsymbol{w} \ne \tilde{\boldsymbol{w}}} \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|. \tag{109}$$

Since $M = \exp(n/30)$, (108) would follow from

$$2c_3\|\boldsymbol{w} - \boldsymbol{x}\|^2 \le n/(300m). \quad \boldsymbol{w} \in \mathcal{M}. \tag{110}$$

Hence, we just need to select δ to be a small multiple of $\sqrt{n/m}$. This in turn gives

$$\inf_{\hat{\boldsymbol{x}}} \sup_{\boldsymbol{x}\in\mathcal{M}} \mathbb{E}\left[\|\hat{\boldsymbol{x}}-\boldsymbol{x}\| \;\middle|\; \{\boldsymbol{a}_i\}\right] \gtrsim \min_{\boldsymbol{w},\tilde{\boldsymbol{w}}\in\mathcal{M},\boldsymbol{w}\neq\tilde{\boldsymbol{w}}} \|\boldsymbol{w}-\tilde{\boldsymbol{w}}\| \gtrsim \sqrt{n/m}. \tag{111}$$

Finally, it remains to connect $\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|$ with $\mathrm{dist}\,(\hat{\boldsymbol{x}},\boldsymbol{x})$. Since all the $\boldsymbol{w}\in\mathcal{M}$ are clustered around $\boldsymbol{x}$ and are at a mutual distance about δ that is much smaller than $\|\boldsymbol{x}\|$, we can see that for any reasonable estimator, $\mathrm{dist}(\hat{\boldsymbol{x}},\boldsymbol{x}) = \|\hat{\boldsymbol{x}}-\boldsymbol{x}\|$. This finishes the proof.

8 Discussion

To keep our treatment concise, this paper does not strive to explore all possible generalizations of the theory. There are nevertheless a few extensions worth pointing out.

- **More general objective functions**. For concreteness, we restrict our analysis to the Poisson log-likelihood function, but the analysis framework we laid out easily carries over to a broad class of (nonconvex) objective functions. For instance, all results continue to hold if we replace the Poisson log-likelihood by the Gaussian log-likelihood; that is, the polynomial function $-\sum_{i=1}^m (y_i - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2)^2$ studied in [1]. A general guideline is to first check whether the expected regularity condition

$$\mathbb{E}\big[-\big\langle \tfrac{1}{m}\nabla \ell_{\mathrm{tr}}\left(\boldsymbol{z}\right), \boldsymbol{h}\big\rangle\big] \;\gtrsim\; \|\boldsymbol{h}\|^2$$

  holds for any fixed $\boldsymbol{z}$ within a neighborhood around $\boldsymbol{x}$. If so, then often times RC holds uniformly within this neighborhood due to sharp concentration of measure ensured by the regularization procedure.

- **Sub-Gaussian measurements**. The theory extends to the situation where the $\boldsymbol{a}_i$'s are i.i.d. sub-Gaussian random vectors, although the truncation threshold might need to be tweaked based on the sub-Gaussian norm of $\boldsymbol{a}_i$. A more challenging scenario, however, is the case where the $\boldsymbol{a}_i$'s are generated according to the CDP model, since there is much less randomness to exploit in the mathematical analysis. We leave this to future research.

Having demonstrated the power of TWF in recovering a rank-one matrix $\boldsymbol{x}\boldsymbol{x}^*$ from quadratic equations, we remark on the potential of TWF towards recovering low-rank matrices from rank-one measurements. Imagine that we wish to estimate a rank-r matrix $\boldsymbol{X}\succeq\boldsymbol{0}$ and that all we know about $\boldsymbol{X}$ is

$$y_i = \boldsymbol{a}_i^\top \boldsymbol{X}\boldsymbol{a}_i, \quad 1\le i\le m.$$

It is known that this problem can be efficiently solved by using more computational-intensive semidefinite programs [9, 14]. With the hope of developing a linear-time algorithm, one might consider a modified TWF scheme, which would maintain a rank-r matrix variable and operate as follows: perform truncated spectral initialization, and then successively update the current guess via a regularized gradient descent rule applied to a presumed log-likelihood function.

Moving away from i.i.d. sub-Gaussian measurements, there is a proliferation of problems that involve completion of a low-rank matrix $\boldsymbol{X}$ from partial entries, where the rank is known *a priori*. It is self-evident that such entry-wise observations can also be cast as rank-one measurements of $\boldsymbol{X}$. Therefore, the preceding modified TWF may add to recent literature in applying non-convex schemes for low-rank matrix completion [42–45], robust PCA [46], or even a broader family of latent-variable models (e.g. dictionary learning [47, 48], sparse coding [49], and mixture problems [50, 51]). A concrete application of this flavor is a simple form of the fundamental alignment/matching problem [52–54]. Imagine a collection of n instances, each representing an image of the same physical object but with different shift $r_i\in\{0,\cdots,M-1\}$. The goal is to align all these instances from observations on the relative shift between pairs of them. Denoting by $\boldsymbol{X}_i$ the cyclic shift by an amount r_i of $\boldsymbol{I}_M$, one sees that the collection matrix $\boldsymbol{X} := [\boldsymbol{X}_i^\top \boldsymbol{X}_j]_{1\le i,j\le k}$ is a rank-M matrix, and the relative shift observations can be treated as rank-one measurements of $\boldsymbol{X}$. Running TWF over this problem instance might result in a statistically and computationally efficient solution. This would be of great practical interest.

A Proofs for Section 5

A.1 Proof of Lemma 3

First, we make the observation that $(\boldsymbol{a}_i^\top \boldsymbol{z})^2 - (\boldsymbol{a}_i^\top \boldsymbol{x})^2 = \left(2\boldsymbol{a}_i^\top \boldsymbol{z} - \boldsymbol{a}_i^\top \boldsymbol{h}\right) \boldsymbol{a}_i^\top \boldsymbol{h}$ is a quadratic function in $\boldsymbol{a}_i^\top \boldsymbol{h}$. If we assume $\gamma \le \frac{\alpha_z^{\mathrm{lb}} \|\boldsymbol{z}\|}{\|\boldsymbol{h}\|}$, then on the event $\mathcal{E}_1^i$ one has

$$(\boldsymbol{a}_i^\top \boldsymbol{z})^2 \;\ge\; \alpha_z^{\mathrm{lb}} \|\boldsymbol{z}\| \cdot |\boldsymbol{a}_i^\top \boldsymbol{z}| \;\ge\; \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|. \tag{112}$$

Solving the quadratic inequality that specifies $\mathcal{D}_\gamma^i$ gives

$$\begin{aligned} \boldsymbol{a}_i^\top \boldsymbol{h} &\in \left[\boldsymbol{a}_i^\top \boldsymbol{z} - \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 + \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|},\;\; \boldsymbol{a}_i^\top \boldsymbol{z} - \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 - \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\right], \\ \text{or} \quad \boldsymbol{a}_i^\top \boldsymbol{h} &\in \left[\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 - \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|},\;\; \boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 + \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\right], \end{aligned}$$

which we will simplify in the sequel.

Suppose for the moment that $\boldsymbol{a}_i^\top \boldsymbol{z} \ge 0$, then the preceding two intervals are respectively equivalent to

$$\begin{aligned} \boldsymbol{a}_i^\top \boldsymbol{h} &\in \underbrace{\left[\frac{-\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 + \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}},\; \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 - \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}}\right]}_{:=I_1}; \\ \boldsymbol{a}_i^\top \boldsymbol{h} - 2\boldsymbol{a}_i^\top \boldsymbol{z} &\in \underbrace{\left[\frac{-\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 - \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}},\; \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 + \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}}\right]}_{:=I_2}. \end{aligned}$$

Assuming (112) and making use of the observations

$$\begin{aligned} \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 - \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}} &\le \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z}} = \gamma \|\boldsymbol{h}\| \\ \text{and} \quad \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\boldsymbol{a}_i^\top \boldsymbol{z} + \sqrt{\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)^2 + \gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}} &\ge \frac{\gamma \|\boldsymbol{h}\| \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}{\left(1+\sqrt{2}\right) \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|} = \frac{\gamma}{1+\sqrt{2}} \|\boldsymbol{h}\|, \end{aligned}$$

we obtain the inner and outer bounds

$$\left[\pm\left(1+\sqrt{2}\right)^{-1} \gamma \|\boldsymbol{h}\|\right] \;\subseteq\; I_1, I_2 \;\subseteq \left[\pm \gamma \|\boldsymbol{h}\|\right].$$

Setting $\gamma_1 := \frac{\gamma}{1+\sqrt{2}}$ gives

$$\left(\mathcal{D}_{\gamma_1}^{i,1} \cap \mathcal{E}_{i,1}\right) \cup \left(\mathcal{D}_{\gamma_1}^{i,2} \cap \mathcal{E}_{i,1}\right) \;\subseteq\; \mathcal{D}_\gamma \cap \mathcal{E}_{i,1} \;\subseteq\; \left(\mathcal{D}_\gamma^{i,1} \cap \mathcal{E}_{i,1}\right) \cup \left(\mathcal{D}_\gamma^{i,2} \cap \mathcal{E}_{i,1}\right).$$

Proceeding with the same argument, we can derive exactly the same inner and outer bounds in the regime where $\boldsymbol{a}_i^\top \boldsymbol{z} < 0$, concluding the proof.

A.2 Proof of Lemma 4

By homogeneity, it suffices to establish the claim for the case where both $\boldsymbol{h}$ and $\boldsymbol{z}$ are *unit vectors*.

Suppose for the moment that $\boldsymbol{h}$ and $\boldsymbol{z}$ are *statistically independent* from $\{\boldsymbol{a}_i\}$. We introduce two auxiliary Lipschitz functions approximating indicator functions:

$$\chi_z(\tau) := \begin{cases} 1, & \text{if } |\tau| \in \left[\sqrt{1.01}\alpha_z^{\mathrm{lb}}, \sqrt{0.99}\alpha_z^{\mathrm{ub}}\right]; \\ -100\left(\alpha_z^{\mathrm{ub}}\right)^{-2}\tau^2 + 100, & \text{if } |\tau| \in \left[\sqrt{0.99}\alpha_z^{\mathrm{ub}}, \alpha_z^{\mathrm{ub}}\right]; \\ 100\left(\alpha_z^{\mathrm{lb}}\right)^{-2}\tau^2 - 100, & \text{if } |\tau| \in \left[\alpha_z^{\mathrm{lb}}, \sqrt{1.01}\alpha_z^{\mathrm{lb}}\right]; \\ 0, & \text{else.} \end{cases} \tag{113}$$

$$\chi_h(\tau) := \begin{cases} 1, & \text{if } |\tau| \in \left[0, \sqrt{0.99}\gamma\right]; \\ -\frac{100}{\gamma^2}\tau^2 + 100, & \text{if } |\tau| \in \left[\sqrt{0.99}\gamma, \gamma\right]; \\ 0, & \text{else.} \end{cases} \tag{114}$$

Since $\boldsymbol{h}$ and $\boldsymbol{z}$ are assumed to be unit vectors, these two functions obey

$$0 \le \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \le \mathbf{1}_{\mathcal{E}_1^i}, \quad \text{and} \quad 0 \le \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) \le \mathbf{1}_{\mathcal{D}_\gamma^{i,1}} \tag{115}$$

and thus,

$$\frac{1}{m}\sum_{i=1}^m \left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1}} \quad \ge \quad \frac{1}{m}\sum_{i=1}^m (\boldsymbol{a}_i^\top \boldsymbol{h})^2 \chi_z(\boldsymbol{a}_i^\top \boldsymbol{z}) \chi_h(\boldsymbol{a}_i^\top \boldsymbol{h}). \tag{116}$$

We proceed to lower bound $\frac{1}{m}\sum_{i=1}^m \left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)$.

Firstly, to compute the mean of $(\boldsymbol{a}_i^\top \boldsymbol{h})^2 \chi_z(\boldsymbol{a}_i^\top \boldsymbol{z}) \chi_h(\boldsymbol{a}_i^\top \boldsymbol{h})$, we introduce an auxiliary orthonormal matrix

$$\boldsymbol{U}_{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{z}^\top / \|\boldsymbol{z}\| \\ \vdots \end{bmatrix} \tag{117}$$

whose first row is along the direction of $\boldsymbol{z}$, and set

$$\tilde{\boldsymbol{h}} := \boldsymbol{U}_{\boldsymbol{z}} \boldsymbol{h}, \quad \text{and} \quad \tilde{\boldsymbol{a}}_i := \boldsymbol{U}_{\boldsymbol{z}} \boldsymbol{a}_i. \tag{118}$$

Also, denote by $\tilde{a}_{i,1}$ (resp. $\tilde{h}_1$) the first entry of $\tilde{\boldsymbol{a}}_i$ (resp. $\tilde{\boldsymbol{h}}$), and $\tilde{\boldsymbol{a}}_{i,\setminus 1}$ (resp. $\tilde{\boldsymbol{h}}_{\setminus 1}$) the remaining entries of $\tilde{\boldsymbol{a}}_i$ (resp. $\tilde{\boldsymbol{h}}$), and let $\xi \sim \mathcal{N}(0,1)$. We have

$$\begin{aligned} &\mathbb{E}\Big[\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) \Big] \\ &\ge \ \mathbb{E}\big[(\boldsymbol{a}_i^\top \boldsymbol{h})^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\big] - \mathbb{E}\left[\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \left(1 - \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)\right)\right] \\ &\ge \ \mathbb{E}\left[\left(\tilde{a}_{i,1}\tilde{h}_1\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\right] + \mathbb{E}\left[\left(\tilde{\boldsymbol{a}}_{i,\setminus 1}^\top \tilde{\boldsymbol{h}}_{\setminus 1}\right)^2\right] \mathbb{E}\left[\chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\right] - \|\boldsymbol{h}\|^2 \mathbb{E}\left[\xi^2 \mathbf{1}_{\left\{|\xi| > \sqrt{0.99}\gamma\right\}}\right] \\ &\ge \ |\tilde{h}_1|^2(1-\zeta_1) + \|\tilde{\boldsymbol{h}}_{\setminus 1}\|^2(1-\zeta_1) - \zeta_2\|\boldsymbol{h}\|^2 \qquad (119) \\ &\ge \ (1-\zeta_1-\zeta_2)\|\boldsymbol{h}\|^2, \end{aligned}$$

where the identity (119) arises from (66) and (67). Since $\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)$ is bounded in magnitude by $\gamma^2\|\boldsymbol{h}\|^2$, it is a sub-Gaussian random variable with sub-Gaussian norm $O(\gamma^2\|\boldsymbol{h}\|^2)$. Apply the Hoeffding-type inequality [40, Proposition 5.10] to deduce that for any $\epsilon > 0$,

$$\frac{1}{m}\sum_{i=1}^m \left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) \ge \mathbb{E}\left[\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)\right] - \epsilon\|\boldsymbol{h}\|^2 \tag{120}$$

$$\ge (1-\zeta_1-\zeta_2-\epsilon)\|\boldsymbol{h}\|^2 \tag{121}$$

with probability at least $1 - \exp(-\Omega(\epsilon^2 m))$.

The next step is to obtain uniform control over all *unit vectors*, for which we adopt a basic version of an ϵ-net argument. Specifically, we construct an ϵ-net $\mathcal{N}_\epsilon$ with cardinality $|\mathcal{N}_\epsilon| \le (1 + 2/\epsilon)^{2n}$ (cf. [40]) such that

for any $(\boldsymbol{h}, \boldsymbol{z})$ with $\|\boldsymbol{h}\| = \|\boldsymbol{z}\| = 1$, there exists a pair $\boldsymbol{h}_0, \boldsymbol{z}_0 \in \mathcal{N}_\epsilon$ satisfying $\|\boldsymbol{h} - \boldsymbol{h}_0\| \le \epsilon$ and $\|\boldsymbol{z} - \boldsymbol{z}_0\| \le \epsilon$. Now that we have discretized the unit spheres using a finite set, taking the union bound gives

$$\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}_0\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right) \ge (1-\zeta_1-\zeta_2-\epsilon)\left\|\boldsymbol{h}_0\right\|^2, \quad \forall \boldsymbol{h}_0, \boldsymbol{z}_0 \in \mathcal{N}_\epsilon \tag{122}$$

with probability at least $1-(1+2/\epsilon)^{2n}\exp(-\Omega(\epsilon^2 m))$.

Define $f_1(\cdot)$ and $f_2(\cdot)$ such that $f_1(\tau) := \tau\chi_h(\sqrt{\tau})$ and $f_2(\tau) := \chi_z(\sqrt{\tau})$, which are both bounded functions with Lipschitz constant $O(1)$. This guarantees that for each *unit* vector pair $\boldsymbol{h}$ and $\boldsymbol{z}$,

$$\begin{aligned}
&\left|\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) - \left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}_0\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)\right| \\
&\le \left|\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\right| \cdot \left|\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) - \left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)^2 \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)\right| \\
&\qquad + \left|(\boldsymbol{a}_i^\top \boldsymbol{h}_0)^2 \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)\right| \cdot \left|\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right) - \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{z}_0\right)\right| \\
&\le \left|\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\right| \cdot \left|f_1\big(|\boldsymbol{a}_i^\top \boldsymbol{h}|^2\big) - f_1\big(|\boldsymbol{a}_i^\top \boldsymbol{h}_0|^2\big)\right| \\
&\qquad + \left|(\boldsymbol{a}_i^\top \boldsymbol{h}_0)^2 \chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)\right| \cdot \left|f_2\left(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2\right) - f_2\left(|\boldsymbol{a}_i^\top \boldsymbol{z}_0|^2\right)\right| \\
&\lesssim \left|(\boldsymbol{a}_i^\top \boldsymbol{h})^2 - (\boldsymbol{a}_i^\top \boldsymbol{h}_0)^2\right| + \left|(\boldsymbol{a}_i^\top \boldsymbol{z})^2 - (\boldsymbol{a}_i^\top \boldsymbol{z}_0)^2\right|.
\end{aligned}$$

Consequently, there exists some universal constant $c_3 > 0$ such that

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) - \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}_0\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}_0\right)\right| \\
&\lesssim \frac{1}{m}\left\|\mathcal{A}\big(\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{h}_0\boldsymbol{h}_0^\top\big)\right\|_1 + \frac{1}{m}\left\|\mathcal{A}\big(\boldsymbol{z}\boldsymbol{z}^\top - \boldsymbol{z}_0\boldsymbol{z}_0^\top\big)\right\|_1 \\
&\overset{\text{(i)}}{\le} c_3\left\{\left\|\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{h}_0\boldsymbol{h}_0^\top\right\|_{\mathrm{F}} + \left\|\boldsymbol{z}\boldsymbol{z}^\top - \boldsymbol{z}_0\boldsymbol{z}_0^\top\right\|_{\mathrm{F}}\right\} \\
&\overset{\text{(ii)}}{\le} 2.5c_3\left\{\left\|\boldsymbol{h} - \boldsymbol{h}_0\right\| \cdot \left\|\boldsymbol{h}\right\| + \left\|\boldsymbol{z} - \boldsymbol{z}_0\right\| \cdot \left\|\boldsymbol{z}\right\|\right\} \le 5c_3\epsilon,
\end{aligned}$$

where (i) results from Lemma 1, and (ii) arises from Lemma 2 whenever $\epsilon < 1/2$.

With the assertion (122) in place, we see that with high probability,

$$\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \chi_z\left(\boldsymbol{a}_i^\top \boldsymbol{z}\right)\chi_h\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right) \quad \ge \quad (1-\zeta_1-\zeta_2-(5c_3+1)\,\epsilon)\left\|\boldsymbol{h}\right\|^2$$

for all unit vectors $\boldsymbol{h}$ and $\boldsymbol{z}$. Since ϵ can be arbitrary, putting this and (116) together completes the proof.

A.3 Proof of Lemma 5

The proof makes use of standard concentration of measure and covering arguments, and it suffices to restrict our attention to *unit vectors* $\boldsymbol{h}$. We find it convenient to work with an auxiliary function

$$\chi_2(\tau) = \begin{cases} |\tau|^{\frac{3}{2}}, & \text{if } |\tau| \le \gamma^2, \\ -\gamma\left(|\tau| - \gamma^2\right) + \gamma^3, & \text{if } \gamma^2 < |\tau| \le 2\gamma^2, \\ 0, & \text{else.} \end{cases}$$

Apparently, $\chi_2(\tau)$ is a Lipschitz function of τ with Lipschitz norm $O(\gamma)$. Recalling the definition of $\mathcal{D}_\gamma^{i,1}$, we see that each summand is bounded above by

$$|\boldsymbol{a}_i^\top \boldsymbol{h}|^3 \, \mathbf{1}_{\mathcal{D}_\gamma^{i,1}} \le \chi_2\big(|\boldsymbol{a}_i^\top \boldsymbol{h}|^2\big).$$

For each fixed $\boldsymbol{h}$ and $\epsilon > 0$, applying the Bernstein inequality [40, Proposition 5.16] gives

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3 \mathbf{1}_{\mathcal{D}_\gamma^{i,1}} &\le \frac{1}{m}\sum_{i=1}^{m}\chi_2\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2\right) \le \mathbb{E}\left[\chi_2\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2\right)\right] + \epsilon \\
&\le \mathbb{E}\big[\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3\big] + \epsilon \;=\; \sqrt{8/\pi} + \epsilon
\end{aligned}$$

with probability exceeding $1-\exp\left(-\Omega\left(\epsilon^2 m\right)\right)$.

From [40, Lemma 5.2], there exists an ϵ-net $\mathcal{N}_\epsilon$ of the unit sphere with cardinality $|\mathcal{N}_\epsilon| \leq \left(1+\frac{2}{\epsilon}\right)^n$. For each $\boldsymbol{h}$, suppose that $\|\boldsymbol{h}_0 - \boldsymbol{h}\| \leq \epsilon$ for some $\boldsymbol{h}_0 \in \mathcal{N}_\epsilon$. The Lipschitz property of χ_2 implies

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m \left\{\chi_2\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2\right) - \chi_2\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2\right)\right\} \;&\lesssim\; \frac{1}{m}\sum_{i=1}^m \left|\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2\right| \\ &\overset{\text{(i)}}{\asymp}\; \|\boldsymbol{h}-\boldsymbol{h}_0\|\,\|\boldsymbol{h}\| \;\asymp\; \epsilon,\end{aligned}$$

where (i) arises by combining Lemmas 1 and 2. This demonstrates that with high probability,

$$\frac{1}{m}\sum_{i=1}^m \left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^3 \mathbf{1}_{\mathcal{D}_\gamma^{i,1}} \leq \frac{1}{m}\sum_{i=1}^m \chi_2\left(|\boldsymbol{a}_i^\top \boldsymbol{h}|^2\right) \leq \sqrt{8/\pi} + O\left(\epsilon\right)$$

for all unit vectors $\boldsymbol{h}$, as claimed.

A.4 Proof of Lemma 6

Without loss of generality, the proof focuses on the case where $\|\boldsymbol{h}\| = 1$. Fix an arbitrary small constant $\delta > 0$. One can eliminate the difficulty of handling the discontinuous indicator functions by working with the following auxiliary function

$$\chi_3\left(\tau,\gamma\right) := \begin{cases} 1, & \text{if } \sqrt{\tau} \geq \psi_{\mathrm{lb}}\left(\gamma\right); \\ \frac{100\tau}{\psi_{\mathrm{lb}}^2(\gamma)} - 99, & \text{if } \sqrt{\tau} \in \left[\sqrt{0.99}\psi_{\mathrm{lb}}\left(\gamma\right),\ \psi_{\mathrm{lb}}\left(\gamma\right)\right]; \\ 0, & \text{else.}\end{cases} \tag{123}$$

Here, $\psi_{\mathrm{lb}}\left(\cdot\right)$ is a piecewise constant function defined as

$$\psi_{\mathrm{lb}}\left(\gamma\right) := (1+\delta)^{\left\lfloor \frac{\log\gamma}{\log(1+\delta)}\right\rfloor},$$

which clearly satisfy $\frac{\gamma}{1+\delta} \leq \psi_{\mathrm{lb}}\left(\gamma\right) \leq \gamma$. Such a function is useful for our purpose since for any $0 < \delta \leq 0.005$,

$$\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq \gamma\right\}} \leq \chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\big) \leq \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq \sqrt{0.99}\psi_{\mathrm{lb}}(\gamma)\right\}} \leq \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}}. \tag{124}$$

For any fixed unit vector $\boldsymbol{h}$, the above argument leads to an upper tail estimate: for any $0 < t \leq 1$,

$$\begin{aligned}\mathbb{P}\left\{\chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\big) \geq t\right\} \;&\leq\; \mathbb{P}\left\{\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}} \geq t\right\} \;=\; \mathbb{P}\left\{\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}} = 1\right\} \\ &=\; 2\int_{0.99\gamma}^{\infty} \phi\left(x\right)\mathrm{d}x \;\leq \frac{2}{0.99\gamma}\phi\left(0.99\gamma\right),\end{aligned} \tag{125}$$

where $\phi(x)$ is the density of a standard normal, and (125) follows from the tail bound $\int_x^\infty \phi(x)\mathrm{d}x \leq \frac{1}{x}\phi\left(x\right)$ for all $x > 0$. This implies that when $\gamma \geq 2$, both $\chi_3\big(|\boldsymbol{a}_i^\top \boldsymbol{h}|^2, \gamma\big)$ and $\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}}$ are sub-exponential with sub-exponential norm $O(\gamma^{-2})$ (cf. [40, Definition 5.13]). We apply the Bernstein-type inequality for the sum of sub-exponential random variables [40, Corollary 5.17], which indicates that for any fixed $\boldsymbol{h}$ and γ as well as any sufficiently small $\epsilon \in (0,1)$,

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m \chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\big) \;&\leq\; \frac{1}{m}\sum_{i=1}^m \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}} \;\leq \mathbb{E}\left[\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \geq 0.99\gamma\right\}}\right] + \epsilon\frac{1}{\gamma^2} \\ &\leq\; \frac{2}{0.99\gamma}\exp\left(-0.49\gamma^2\right) + \epsilon\frac{1}{\gamma^2}\end{aligned}$$

holds with probability exceeding $1-\exp\left(-\Omega(\epsilon^2 m)\right)$.

We now proceed to obtain uniform control over all $\boldsymbol{h}$ and $2 \le \gamma \le 2^n$. To begin with, we consider all $2 \le \gamma \le m$ and construct an ϵ-net $\mathcal{N}_\epsilon$ over the unit sphere such that: (i) $|\mathcal{N}_\epsilon| \le \left(1+\frac{2}{\epsilon}\right)^n$; (ii) for any $\boldsymbol{h}$ with $\|\boldsymbol{h}\| = 1$, there exists a unit vector $\boldsymbol{h}_0 \in \mathcal{N}_\epsilon$ obeying $\|\boldsymbol{h} - \boldsymbol{h}_0\| \le \epsilon$. Taking the union bound gives the following: with probability at least $1 - \frac{\log m}{\log(1+\delta)}\left(1+\frac{2}{\epsilon}\right)^n \exp(-\Omega(\epsilon^2 m))$,

$$\frac{1}{m}\sum_{i=1}^m \chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2,\ \gamma_0\big) \quad \le \quad (0.495\gamma_0)^{-1}\exp\left(-0.49\gamma_0^2\right) + \epsilon\gamma_0^{-2}$$

holds simultaneously for all $\boldsymbol{h}_0 \in \mathcal{N}_\epsilon$ and $\gamma_0 \in \left\{(1+\delta)^k \mid 1 \le k \le \frac{\log m}{\log(1+\delta)}\right\}$.

Note that $\chi_3(\tau, \gamma_0)$ is a Lipschitz function in τ with the Lipschitz constant bounded above by $\frac{100}{\psi_{\rm lb}^2(\gamma_0)}$. With this in mind, for any $(\boldsymbol{h}, \gamma)$ with $\|\boldsymbol{h}\| = 1$ and $\gamma_0 := (1+\delta)^k \le \gamma < (1+\delta)^{k+1}$, one has

$$\begin{aligned}\left|\chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2,\ \gamma_0\big) - \chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\big)\right| &= \left|\chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2,\ \gamma_0\big) - \chi_3\big(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma_0\big)\right| \\ &\le \frac{100}{\psi_{\rm lb}^2(\gamma_0)}\left|\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2 - \left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2\right|.\end{aligned}$$

It then follows from Lemmas 1-2 that

$$\begin{aligned}&\frac{1}{m}\left|\sum_{i=1}^m \chi_3\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2,\ \gamma_0\right) - \sum_{i=1}^m \chi_3\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\right)\right| \ \le\ \frac{100}{\psi_{\rm lb}^2(\gamma_0)}\frac{1}{m}\left\|\mathcal{A}\left(\boldsymbol{h}\boldsymbol{h}^\top - \boldsymbol{h}_0\boldsymbol{h}_0^\top\right)\right\|_1 \\ &\le \frac{250(1+\delta)^2}{\gamma^2}\|\boldsymbol{h}-\boldsymbol{h}_0\|\|\boldsymbol{h}\| \ \le\ \frac{250(1+\delta)^2\epsilon}{\gamma^2}.\end{aligned}$$

Putting the above results together gives that for all $2 \le \gamma \le (1+\delta)^{\frac{\log m}{\log(1+\delta)}} = m$,

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m \chi_3\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2,\ \gamma\right) &\le \frac{1}{m}\sum_{i=1}^m \chi_3\left(\left|\boldsymbol{a}_i^\top \boldsymbol{h}_0\right|^2,\ \gamma_0\right) + \frac{250(1+\delta)^2}{\gamma^2}\epsilon \\ &\le \frac{1}{0.495\gamma_0}\exp\left(-0.49\gamma_0^2\right) + 251(1+\delta)^2\frac{\epsilon}{\gamma^2} \\ &\le \frac{1}{0.49\gamma}\exp\left(-0.485\gamma^2\right) + 251(1+\delta)^2\frac{\epsilon}{\gamma^2}\end{aligned}$$

with probability exceeding $1 - \frac{\log m}{\log(1+\delta)}\left(1+\frac{2}{\epsilon}\right)^n \exp\left(-c\epsilon^2 m\right)$. This establishes (71) for all $2 \le \gamma \le m$.

It remains to deal with the case where $\gamma > m$. To this end, we rely on the following observation:

$$\frac{1}{m}\sum_{i=1}^m \mathbf{1}_{\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \ge m\}} \ \le\ \frac{1}{m}\sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2}{m^2} \ \overset{\text{(i)}}{\le}\ \frac{1+\delta}{m^2}\|\boldsymbol{h}\|^2 \ \ll\ \frac{1}{m}, \quad \forall \boldsymbol{h} \text{ with } \|\boldsymbol{h}\| = 1,$$

where (i) comes from [6, Lemmas 3.1]. This basically tells us that with high probability, none of the indicator variables can be equal to 1. Consequently, $\frac{1}{m}\sum_{i=1}^m \mathbf{1}_{\{|\boldsymbol{a}_i^\top \boldsymbol{h}| \ge m\}} = 0$, which proves the claim.

A.5 Proof of Lemma 7

Fix $\delta > 0$. Recalling the notation $v_i := 2\left\{2\boldsymbol{a}_i^\top \boldsymbol{h} - \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}\right\}\mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i}$, we see from the expansion (62) that

$$\left\|\frac{1}{m}\nabla_{\rm tr}\ell(\boldsymbol{z})\right\| = \left\|\frac{1}{m}\boldsymbol{A}^\top \boldsymbol{v}\right\| \le \frac{1}{m}\|\boldsymbol{A}\| \cdot \|\boldsymbol{v}\| \le (1+\delta)\frac{\|\boldsymbol{v}\|}{\sqrt{m}} \tag{126}$$

as soon as $m \ge c_1 n$ for some sufficiently large $c_1 > 0$. Here, the norm estimate $\|\boldsymbol{A}\| \le \sqrt{m}(1+\delta)$ arises from standard random matrix results [40, Corollary 5.35].

Everything then comes down to controlling $\|\boldsymbol{v}\|$. To this end, making use of the inclusion (63) yields

$$\begin{aligned}
\frac{1}{4m}\|\boldsymbol{v}\|^2 &= \frac{1}{m}\sum_{i=1}^m \left(2\boldsymbol{a}_i^\top \boldsymbol{h} - \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^2}{\boldsymbol{a}_i^\top \boldsymbol{z}}\right)^2 \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} \\
&\le \frac{1}{m}\sum_{i=1}^m \left(2\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right| + \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^2}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\right)^2 \mathbf{1}_{\mathcal{E}_1^i \cap \left(\mathcal{D}_{\gamma_4}^{i,1} \cup \mathcal{D}_{\gamma_4}^{i,2}\right)} \\
&\le \frac{1}{m}\sum_{i=1}^m \left\{4(\boldsymbol{a}_i^\top \boldsymbol{h})^2 + \left(\frac{4|\boldsymbol{a}_i^\top \boldsymbol{h}|^3}{|\boldsymbol{a}_i^\top \boldsymbol{z}|} + \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^4}{|\boldsymbol{a}_i^\top \boldsymbol{z}|^2}\right) \mathbf{1}_{\mathcal{E}_1^i \cap \left(\mathcal{D}_{\gamma_4}^{i,1} \cup \mathcal{D}_{\gamma_4}^{i,2}\right)}\right\} \\
&= \frac{1}{m}\sum_{i=1}^m \left\{4\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 + \left(4 + \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\right)\frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^3}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\left(\mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_4}^{i,1}} + \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_4}^{i,2}}\right)\right\}.
\end{aligned}$$

The first term is controlled by [6, Lemma 3.1] in such a way that with probability $1 - \exp(-\Omega(m))$,

$$\frac{1}{m}\sum_{i=1}^m 4\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 \le 4(1+\delta)\|\boldsymbol{h}\|^2.$$

Turning to the remaining terms, we see from the definition of $\mathcal{D}_\gamma^{i,1}$ and $\mathcal{D}_\gamma^{i,2}$ that

$$\frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|}{|\boldsymbol{a}_i^\top \boldsymbol{z}|} \le \begin{cases} \frac{\gamma\|\boldsymbol{h}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}, & \text{on } \mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1} \\ 2 + \frac{\gamma\|\boldsymbol{h}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}, & \text{on } \mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,2} \end{cases} \le \begin{cases} 1, & \text{on } \mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1} \\ 3, & \text{on } \mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,2} \end{cases}$$

as long as $\gamma \le \frac{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}{\|\boldsymbol{h}\|}$. Consequently, one can bound

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^m \left(4 + \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\right)\frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^3}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\left(\mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1}} + \mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,2}}\right) \\
&\le \frac{5}{m}\sum_{i=1}^m \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^3}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,1}} + \frac{7}{m}\sum_{i=1}^m \frac{|\boldsymbol{a}_i^\top \boldsymbol{h}|^3}{|\boldsymbol{a}_i^\top \boldsymbol{z}|}\mathbf{1}_{\mathcal{E}_1^i \cap \mathcal{D}_\gamma^{i,2}} \\
&\le \frac{5(1+\delta)\sqrt{8/\pi}\|\boldsymbol{h}\|^3}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|} + \frac{7}{100}(1+\delta)\|\boldsymbol{h}\|^2,
\end{aligned}$$

where the last inequality follows from (69) and (78).

Recall that $\gamma_4 = 3\alpha_h$. Taken together all these bounds lead to the upper bound

$$\begin{aligned}
\frac{1}{4m}\|\boldsymbol{v}\|^2 &\le (1+\delta)\left\{4 + \frac{5\sqrt{8/\pi}\,\|\boldsymbol{h}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|} + \frac{7}{100}\right\}\|\boldsymbol{h}\|^2 \\
&\le (1+\delta)\left\{4 + \frac{5\sqrt{8/\pi}}{3\alpha_h} + \frac{7}{100}\right\}\|\boldsymbol{h}\|^2
\end{aligned}$$

whenever $\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \min\left\{\frac{\alpha_z^{\mathrm{lb}}}{3\alpha_h}, \frac{\alpha_z^{\mathrm{lb}}}{6}, \frac{\sqrt{98/3}\left(\alpha_z^{\mathrm{lb}}\right)^2}{2\alpha_z^{\mathrm{ub}} + \alpha_z^{\mathrm{lb}}}, \frac{1}{11}\right\}$. Substituting this into (126) completes the proof.

B Proofs for Section 7

B.1 Proof of Lemma 8

Firstly, we collect a few results on the magnitudes of $\boldsymbol{a}_i^\top \boldsymbol{x}$ $(1 \le i \le m)$ that will be useful in constructing the hypotheses. Observe that for any given $\boldsymbol{x}$ and any sufficiently large m,

$$\begin{aligned}
\mathbb{P}\left\{\min_{1\le i\le m}\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right| \ge \frac{1}{m\log m}\|\boldsymbol{x}\|\right\} &= \left(\mathbb{P}\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \ge \frac{1}{m\log m}\|\boldsymbol{x}\|\right\}\right)^m \\
&\ge \left(1 - \frac{2}{\sqrt{2\pi}}\frac{1}{m\log m}\right)^m \ge 1 - o(1).
\end{aligned}$$

Besides, since $\mathbb{E}\left[\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|\le \frac{\|\boldsymbol{x}\|}{5\log m}\right\}}\right] \le \frac{1}{\sqrt{2\pi}}\frac{2}{5\log m} \le \frac{1}{5\log m}$, applying Hoeffding's inequality yields

$$\begin{aligned}
&\mathbb{P}\bigg\{\sum\nolimits_{i=1}^m \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|\le \frac{\|\boldsymbol{x}\|}{5\log m}\right\}} > \frac{m}{4\log m}\bigg\} \\
&= \mathbb{P}\bigg\{\frac{1}{m}\sum\nolimits_{i=1}^m \Big(\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|\le \frac{\|\boldsymbol{x}\|}{5\log m}\right\}} - \mathbb{E}\left[\mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|\le \frac{\|\boldsymbol{x}\|}{5\log m}\right\}}\right]\Big) > \frac{1}{20\log m}\bigg\} \\
&\le \exp\left(-\Omega\Big(\frac{m}{\log^2 m}\Big)\right).
\end{aligned}$$

To summarize, with probability $1-o(1)$, one has

$$\min\nolimits_{1\le i\le m} \left|\boldsymbol{a}_i^\top \boldsymbol{x}\right| \;\ge\; \frac{1}{m\log m}\|\boldsymbol{x}\|; \tag{127}$$

$$\sum\nolimits_{i=1}^m \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|\le \frac{\|\boldsymbol{x}\|}{\log m}\right\}} \;\le\; \frac{m}{4\log m} := k. \tag{128}$$

In the sequel, we will first produce a set $\mathcal{M}_1$ of exponentially many vectors surrounding $\boldsymbol{x}$ in such a way that every pair is separated by about the same distance, and then verify that a non-trivial fraction of $\mathcal{M}_1$ obeys (103). Without loss of generality, we assume that $\boldsymbol{x}$ takes the form $\boldsymbol{x} = [b, 0, \cdots, 0]^\top$ for some $b > 0$.

The construction of $\mathcal{M}_1$ follows a standard random packing argument. Let $\boldsymbol{w} = [w_1, \cdots, w_n]^\top$ be a random vector with

$$w_i = x_i + \frac{1}{\sqrt{2n}} z_i, \quad 1 \le i \le n,$$

where $z_i \overset{\text{ind.}}{\sim} \mathcal{N}(0,1)$. The collection $\mathcal{M}_1$ is then obtained by generating $M_1 = \exp\left(\frac{n}{20}\right)$ independent copies $\boldsymbol{w}^{(l)}$ $(1 \le l < M_1)$ of $\boldsymbol{w}$. For any $\boldsymbol{w}^{(l)}, \boldsymbol{w}^{(j)} \in \mathcal{M}_1$, the concentration inequality [40, Corollary 5.35] gives

$$\begin{aligned}
\mathbb{P}\left\{0.5\sqrt{n} - 1 \le \sqrt{n}\left\|\boldsymbol{w}^{(l)} - \boldsymbol{w}^{(j)}\right\| \le 1.5\sqrt{n} + 1\right\} &\ge 1 - 2\exp\left(-n/8\right); \\
\mathbb{P}\left\{0.5\sqrt{n} - 1 \le \sqrt{2n}\left\|\boldsymbol{w}^{(l)} - \boldsymbol{x}\right\| \le 1.5\sqrt{n} + 1\right\} &\ge 1 - 2\exp\left(-n/8\right).
\end{aligned}$$

Taking the union bound over all $\binom{M_1}{2}$ pairs we obtain

$$\begin{aligned}
0.5 - n^{-1/2} \le \;& \left\|\boldsymbol{w}^{(l)} - \boldsymbol{w}^{(j)}\right\| \;\le 1.5 + n^{-1/2}, \quad \forall l \ne j \\
1/\sqrt{8} - (2n)^{-1/2} \le \;& \left\|\boldsymbol{w}^{(l)} - \boldsymbol{x}\right\| \;\le \sqrt{9/8} + (2n)^{-1/2}, \quad 1 \le l \le M_1
\end{aligned} \tag{129}$$

with probability exceeding $1 - 2M_1^2\exp\left(-\frac{n}{8}\right) \ge 1 - 2\exp\left(-\frac{n}{40}\right)$.

The next step is to show that many vectors in $\mathcal{M}_1$ satisfy (103). For any given $\boldsymbol{w}$ with $\boldsymbol{r} := \boldsymbol{w} - \boldsymbol{x}$, by letting $\boldsymbol{a}_{i,\perp} := [a_{i,2}, \cdots, a_{i,n}]^\top$, $r_\parallel := r_1$, and $\boldsymbol{r}_\perp := [r_2, \cdots, r_n]^\top$, we derive

$$\frac{|\boldsymbol{a}_i^\top \boldsymbol{r}|^2}{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2} \le \frac{2|a_{i,1} r_\parallel|^2 + 2|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp|^2}{|a_{i,1}|^2 \|\boldsymbol{x}\|^2} \le \frac{2|r_\parallel|^2}{\|\boldsymbol{x}\|^2} + \frac{2|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp|^2}{|a_{i,1}|^2 \|\boldsymbol{x}\|^2} \le \frac{2\|\boldsymbol{r}\|^2}{\|\boldsymbol{x}\|^2} + \frac{2|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp|^2}{|a_{i,1}|^2 \|\boldsymbol{x}\|^2}. \tag{130}$$

It then boils down to developing an upper bound on $\frac{\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp\right|^2}{|a_{i,1}|^2}$. This ratio is convenient to work with since the numerator and denominator are stochastically independent. To simplify presentation, we reorder $\{\boldsymbol{a}_i\}$ in a way that

$$(m\log m)^{-1}\|\boldsymbol{x}\| \le \left|\boldsymbol{a}_1^\top \boldsymbol{x}\right| \le \left|\boldsymbol{a}_2^\top \boldsymbol{x}\right| \le \cdots \le \left|\boldsymbol{a}_m^\top \boldsymbol{x}\right|;$$

this will not affect our subsequent analysis concerning $\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp$ since it is independent of $\boldsymbol{a}_i^\top \boldsymbol{x}$.

To proceed, we let $\boldsymbol{r}_\perp^{(l)}$ consist of all but the first entry of $\boldsymbol{w}^{(l)} - \boldsymbol{x}$, and introduce the indicator variables

$$\xi_i^l := \begin{cases} \mathbf{1}_{\left\{\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| \le \frac{1}{m}\sqrt{\frac{n-1}{2n}}\right\}}, & 1 \le i \le k, \\ \mathbf{1}_{\left\{\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| \le \sqrt{\frac{2(n-1)\log n}{n}}\right\}}, & i > k, \end{cases} \tag{131}$$

where $k = \frac{m}{4\log m}$ as before. In words, we divide $\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}$, $1 \le i \le m$ into two groups, with the first group enforcing far more stringent control than the second group. These indicator variables are useful since any $\boldsymbol{w}^{(l)}$ obeying $\prod_{i=1}^m \xi_i^l = 1$ will satisfy (103) when n is sufficiently large. To see this, note that for the first group of indices, $\xi_i^l = 1$ requires

$$\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| \le \frac{1}{m}\sqrt{\frac{n-1}{2n}} \le \frac{2}{m}\frac{\sqrt{n-1}}{\sqrt{n}-2}\left\|\boldsymbol{r}^{(l)}\right\| \le \frac{3}{m}\left\|\boldsymbol{r}^{(l)}\right\|, \quad 1 \le i \le k, \tag{132}$$

where the second inequality follows from (129). This taken collectively with (127) and (130) yields

$$\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{r}^{(l)}\right|^2}{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2} \le \frac{2\|\boldsymbol{r}^{(l)}\|^2}{\|\boldsymbol{x}\|^2} + \frac{\frac{9}{m^2}\left\|\boldsymbol{r}^{(l)}\right\|^2}{\frac{1}{m^2\log^2 m}\|\boldsymbol{x}\|^2} \le \frac{(2+9\log^2 m)\left\|\boldsymbol{r}^{(l)}\right\|^2}{\|\boldsymbol{x}\|^2}, \quad 1 \le i \le k.$$

Regarding the second group of indices, $\xi_i^l = 1$ gives

$$\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| \le \sqrt{\frac{2\,(n-1)\log n}{n}} \le \sqrt{17\log n}\left\|\boldsymbol{r}^{(l)}\right\|, \quad i = k+1, \cdots, m, \tag{133}$$

where the last inequality again follows from (129). Plugging (133) and (128) into (130) gives

$$\frac{\left|\boldsymbol{a}_i^\top \boldsymbol{r}^{(l)}\right|^2}{\left|\boldsymbol{a}_i^\top \boldsymbol{x}\right|^2} \le \frac{2\|\boldsymbol{r}^{(l)}\|^2}{\|\boldsymbol{x}\|^2} + \frac{17\left\|\boldsymbol{r}^{(l)}\right\|^2\log n}{\|\boldsymbol{x}\|^2/\log^2 m} \le \frac{(2+17\log^3 m)\left\|\boldsymbol{r}^{(l)}\right\|^2}{\|\boldsymbol{x}\|^2}, \quad i \ge k+1.$$

Consequently, (103) is satified for all $1 \le i \le m$. It then suffices to guarantee the existence of exponentially many vectors obeying $\prod_{i=1}^m \xi_i^l = 1$.

Note that the first group of indicator variables are quite stringent, namely, for each i only a fraction $O(1/m)$ of the equations could satisfy $\xi_i^l = 1$. Fortunately, M_1 is exponentially large, and hence even M_1/m^k is exponentially large. Put formally, we claim that the first group satisfies

$$\sum_{l=1}^{M_1}\prod_{i=1}^{k}\xi_i^l \ge \frac{1}{2}\frac{M_1}{(2\pi)^{k/2}\,(1+4\sqrt{k/n})^{k/2}}\left(\frac{1}{\sqrt{2\pi}m}\right)^k := \widetilde{M}_1 \tag{134}$$

with probability exceeding $1 - \exp\left(-\Omega\left(k\right)\right) - \exp(-\widetilde{M}_1/4)$. With this claim in place (which will be proved later), one has

$$\sum_{l=1}^{M_1}\prod_{i=1}^{k}\xi_i^l \ge \frac{1}{2}M_1\frac{1}{(e^2m)^k} = \frac{1}{2}\exp\left(\left(\frac{1}{20} - \frac{k\,(2+\log m)}{n}\right)n\right) \ge \frac{1}{2}\exp\left(\frac{1}{25}n\right)$$

when n and m/n are sufficiently large. In light of this, we will let $\mathcal{M}_2$ be a collection comprising all $\boldsymbol{w}^{(l)}$ obeying $\prod_{i=1}^k \xi_i^l = 1$, which has size $M_2 \ge \frac{1}{2}\exp\left(\frac{1}{25}n\right)$ based on the preceding argument. For notational simplicity, it will be assumed that the vectors in $\mathcal{M}_2$ are exactly $\boldsymbol{w}^{(j)}$ $(1 \le j \le M_2)$.

We now move on to the second group by examining how many vectors $\boldsymbol{w}^{(j)}$ in $\mathcal{M}_2$ further satisfy $\prod_{i=k+1}^m \xi_i^j = 1$. Notably, the above construction of $\mathcal{M}_2$ relies only on $\{\boldsymbol{a}_i\}_{1\le i\le k}$ and is independent of the remaining vectors $\{\boldsymbol{a}_i\}_{i>k}$. In what follows the argument proceeds conditional on $\mathcal{M}_2$ and $\{\boldsymbol{a}_i\}_{1\le i\le k}$. Applying the union bound gives

$$\begin{aligned}
&\mathbb{E}\left[\sum\nolimits_{j=1}^{M_2}\left(1-\prod\nolimits_{i=k+1}^{m}\xi_i^j\right)\right] = \sum\nolimits_{j=1}^{M_2}\mathbb{P}\left\{\exists i\ (k<i\le m):\ \left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| > \sqrt{\frac{2\,(n-1)\log n}{n}}\right\} \\
&\le \sum_{j=1}^{M_2}\sum_{i=k+1}^{m}\mathbb{P}\left\{\left|\boldsymbol{a}_{i,\perp}^\top \boldsymbol{r}_\perp^{(l)}\right| > \sqrt{\frac{2\,(n-1)\log n}{n}}\right\} \le M_2 m\frac{1}{n^2}.
\end{aligned}$$

This combined with Markov's inequality gives

$$\sum\nolimits_{j=1}^{M_2}\left(1-\prod\nolimits_{i=k+1}^{m}\xi_i^j\right) \le \frac{m\log m}{n^2}\cdot M_2$$

with probability $1-o(1)$. Putting the above inequalities together suggests that with probability $1-o(1)$, there exist at least

$$\left(1-\frac{m\log m}{n^2}\right)M_2 \geq \frac{1}{2}\left(1-\frac{m\log m}{n^2}\right)\exp\left(\frac{1}{25}n\right) \geq \exp\left(\frac{n}{30}\right)$$

vectors in $\mathcal{M}_2$ satisfying $\prod_{l=k+1}^{m}\xi_i^l = 1$. We then choose $\mathcal{M}$ to be the set consisting of all these vectors, which forms a valid collection satisfying the properties of Lemma 8.

Finally, the only remaining step is to establish the claim (134). To start with, consider an $n\times k$ matrix $\boldsymbol{B} := [\boldsymbol{b}_1,\cdots,\boldsymbol{b}_k]$ of i.i.d. standard normal entries, and let $\boldsymbol{u}\sim\mathcal{N}\left(\boldsymbol{0},\frac{1}{n}\boldsymbol{I}_n\right)$. Conditional on the $\{\boldsymbol{b}_i$'s,

$$\boldsymbol{b}_{\boldsymbol{u}} = \begin{bmatrix} b_{1,\boldsymbol{u}} \\ \vdots \\ b_{k,\boldsymbol{u}} \end{bmatrix} := \begin{bmatrix} \boldsymbol{b}_1^\top\boldsymbol{u} \\ \vdots \\ \boldsymbol{b}_k^\top\boldsymbol{u} \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0},\ \frac{1}{n}\boldsymbol{B}^\top\boldsymbol{B}\right).$$

For sufficiently large m, one has $k=\frac{m}{4\log m}\leq\frac{1}{4}n$. Using [40, Corollary 5.35] we get

$$\left\|\frac{1}{n}\boldsymbol{B}^\top\boldsymbol{B}-\boldsymbol{I}\right\| \leq 4\sqrt{k/n} \tag{135}$$

with probability $1-\exp\left(-\Omega(k)\right)$. Thus, for any constant $0<\epsilon<\frac{1}{2}$, conditional on $\{\boldsymbol{b}_i\}$ and (135) we obtain

$$\begin{aligned}
&\mathbb{P}\left\{\bigcap_{i=1}^{k}\left\{|\boldsymbol{b}_i^\top\boldsymbol{u}|\leq\frac{1}{m}\right\}\right\} && (136)\\
&\geq\ (2\pi)^{-\frac{k}{2}}\det{}^{-\frac{1}{2}}\Big(\frac{1}{n}\boldsymbol{B}^\top\boldsymbol{B}\Big)\int_{\boldsymbol{b}_{\boldsymbol{u}}\in\Upsilon}\exp\Big(-\frac{1}{2}\boldsymbol{b}_{\boldsymbol{u}}^\top\Big(\frac{1}{n}\boldsymbol{B}^\top\boldsymbol{B}\Big)^{-1}\boldsymbol{b}_{\boldsymbol{u}}\Big)\mathrm{d}\boldsymbol{b}_{\boldsymbol{u}}\\
&\geq\ (2\pi)^{-\frac{k}{2}}\left(1+4\sqrt{k/n}\right)^{-\frac{k}{2}}\int_{\boldsymbol{b}_{\boldsymbol{u}}\in\Upsilon}\exp\Big(-\frac{1}{2}\left(1-4\sqrt{k/n}\right)^{-1}\sum_{i=1}^{k}b_{i,u}^2\Big)\mathrm{d}\boldsymbol{b}_{\boldsymbol{u}} && (137)\\
&\geq\ (2\pi)^{-\frac{k}{2}}\left(1+4\sqrt{k/n}\right)^{-\frac{k}{2}}\left(\sqrt{2\pi}m\right)^{-k}, && (138)
\end{aligned}$$

where $\Upsilon := \{\tilde{\boldsymbol{b}}\mid|\tilde{b}_i|\leq m^{-1}, 1\leq i\leq k\}$ and (137) is a direct consequence from (135).

When it comes to our quantity of interest, the above lower bound (138) indicates that on an event (defined via $\{\boldsymbol{a}_i\}$) of probability approaching 1, we have

$$\mathbb{E}\Big[\sum\nolimits_{l=1}^{M_1}\prod\nolimits_{i=1}^{k}\xi_i^l\Big] \quad\geq\quad M_1\,(2\pi)^{-\frac{k}{2}}\left(1+4\sqrt{k/n}\right)^{-\frac{k}{2}}\left(\sqrt{2\pi}m\right)^{-k}. \tag{139}$$

Since conditional on $\{\boldsymbol{a}_i\}$, $\prod_{i=1}^{k}\xi_i^l$ are independent across l, applying the Chernoff-type bound [55, Theorem 4.5] gives

$$\sum\nolimits_{l=1}^{M_1}\prod\nolimits_{i=1}^{k}\xi_i^l \geq \frac{M_1}{2}\,(2\pi)^{-\frac{k}{2}}\left(1+4\sqrt{k/n}\right)^{-\frac{k}{2}}\left(\sqrt{2\pi}m\right)^{-k}$$

with probability exceeding $1-\exp\Big(-\frac{1}{8}\frac{M_1}{(2\pi)^{k/2}(1+4\sqrt{k/n})^{k/2}}\Big(\frac{1}{\sqrt{2\pi}m}\Big)^k\Big)$. This concludes the proof.

B.2 Proof of Lemma 9

Before proceeding, we introduce the χ^2-divergence between two probability measures P and Q as

$$\chi^2\left(P\|Q\right) := \int\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^2\mathrm{d}Q - 1. \tag{140}$$

It is well known (e.g. [41, Lemma 2.7]) that

$$\mathsf{KL}\left(P\|Q\right) \leq \log(1+\chi^2\left(P\|Q\right)), \tag{141}$$

and hence it suffices to develop an upper bound on the χ^2 divergence.

Under independence, for any $\boldsymbol{w}_0, \boldsymbol{w}_1 \in \mathbb{R}^n$, the decoupling identity of the χ^2 divergence [41, Page 96] gives

$$\begin{aligned}\chi^2\left(\mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{w}_1\right) \parallel \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{w}_0\right)\right) &= \prod_{i=1}^m \left(1+\chi^2\left(\mathbb{P}\left(y_i \mid \boldsymbol{w}_1\right) \parallel \mathbb{P}\left(y_i \mid \boldsymbol{w}_0\right)\right)\right)-1 \\ &= \exp\left(\sum_{i=1}^m \frac{\left(|\boldsymbol{a}_i^\top \boldsymbol{w}_1|^2-|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2\right)^2}{|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2}\right)-1. \end{aligned} \tag{142}$$

The preceding identity (142) arises from the following computation: by definition of $\chi^2(\cdot\|\cdot)$,

$$\begin{aligned}&\chi^2\left(\mathsf{Poisson}\left(\lambda_1\right) \parallel \mathsf{Poisson}\left(\lambda_0\right)\right)=\left\{\sum_{k=0}^\infty \frac{\left(\lambda_1^k \exp\left(-\lambda_1\right)\right)^2}{\lambda_0^k \exp\left(-\lambda_0\right) k!}\right\}-1 \\ &=\exp\Big(\lambda_0-2\lambda_1+\frac{\lambda_1^2}{\lambda_0}\Big)\Big\{\sum_{k=0}^\infty \frac{\left(\lambda_1^2/\lambda_0\right)^k}{k!}\exp\Big(-\frac{\lambda_1^2}{\lambda_0}\Big)\Big\}-1=\exp\Big(\frac{\left(\lambda_1-\lambda_0\right)^2}{\lambda_0}\Big)-1.\end{aligned}$$

Set $\boldsymbol{r} := \boldsymbol{w}_1 - \boldsymbol{w}_0$. To summarize,

$$\begin{aligned}\mathsf{KL}\left(\mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{w}_1\right) \parallel \mathbb{P}\left(\boldsymbol{y} \mid \boldsymbol{w}_0\right)\right) &\le \sum_{i=1}^m \frac{\left(|\boldsymbol{a}_i^\top \boldsymbol{w}_1|^2-|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2\right)^2}{|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2} \qquad (143)\\ &\le \sum_{i=1}^m \frac{\left|\boldsymbol{a}_i^\top \boldsymbol{r}\right|^2\left(2\left|\boldsymbol{a}_i^\top \boldsymbol{w}_0\right|+\left|\boldsymbol{a}_i^\top \boldsymbol{r}\right|\right)^2}{|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2} \\ &= \sum_{i=1}^m |\boldsymbol{a}_i^\top \boldsymbol{r}|^2\left(\frac{8|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2+2|\boldsymbol{a}_i^\top \boldsymbol{r}|^2}{|\boldsymbol{a}_i^\top \boldsymbol{w}_0|^2}\right). \qquad (144)\end{aligned}$$

C Initialization via truncated spectral Method

This section demonstrates that the truncated spectral method works when $m \asymp n$, as stated in the proposition below.

Proposition 3. *Fix $\delta > 0$ and $\boldsymbol{x} \in \mathbb{R}^n$. Consider the model where $y_i = |\boldsymbol{a}_i^\top \boldsymbol{x}|^2 + \eta_i$ and $\boldsymbol{a}_i \overset{\text{ind.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Suppose that*

$$|\eta_i| \le \varepsilon \max\{\|\boldsymbol{x}\|^2,\ |\boldsymbol{a}_i^\top \boldsymbol{x}|^2\}, \quad 1 \le i \le m \tag{145}$$

for some sufficiently small constant $\varepsilon > 0$. With probability exceeding $1 - \exp(-\Omega(m))$, the solution $\boldsymbol{z}^{(0)}$ returned by the truncated spectral method obeys

$$\operatorname{dist}(\boldsymbol{z}^{(0)}, \boldsymbol{x}) \le \delta \|\boldsymbol{x}\|, \tag{146}$$

provided that $m > c_0 n$ for some constant $c_0 > 0$.

Proof. By homogeneity, it suffices to consider the case where $\|\boldsymbol{x}\| = 1$. Recall from [6, Lemma 3.1] that $\frac{1}{m}\sum_{i=1}^m |\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \in [1 \pm \varepsilon]\|\boldsymbol{x}\|^2$ holds with probability $1 - \exp(-\Omega(m))$. Under the hypothesis (145), one has

$$\frac{1}{m}\|\boldsymbol{\eta}\|_1 \le \frac{1}{m}\sum_{i=1}^m \varepsilon(\|\boldsymbol{x}\|^2 + |\boldsymbol{a}_i^\top \boldsymbol{x}|^2) \le \varepsilon\|\boldsymbol{x}\|^2 + \varepsilon(1+\epsilon)\|\boldsymbol{x}\|^2 \le 3\varepsilon\|\boldsymbol{x}\|^2,$$

which yields

$$\lambda_0^2 := \frac{1}{m}\sum_{l=1}^m y_l \ = \ \frac{1}{m}\sum_{l=1}^m |\boldsymbol{a}_l^\top \boldsymbol{x}|^2 + \frac{1}{m}\sum_{l=1}^m \eta_l \ \in \ [1 \pm 4\varepsilon]\|\boldsymbol{x}\|^2 \tag{147}$$

with probability $1 - \exp(-\Omega(m))$.

Consequently, when $|\eta_i| \le \varepsilon \|\boldsymbol{x}\|^2$, one has

$$\begin{array}{l} \mathbf{1}_{\left\{|(\boldsymbol{a}_i^\top \boldsymbol{x})^2+\eta_i| \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \le \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \alpha_y^2(\frac{1}{m}\sum_l y_l)+|\eta_i|\right\}} \le \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le (1+4\varepsilon)\alpha_y^2+\varepsilon\right\}}; \\ \mathbf{1}_{\left\{|(\boldsymbol{a}_i^\top \boldsymbol{x})^2+\eta_i| \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \ge \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \alpha_y^2(\frac{1}{m}\sum_l y_l)-|\eta_i|\right\}} \ge \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le (1-4\varepsilon)\alpha_y^2-\varepsilon\right\}}. \end{array} \tag{148}$$

Besides, in the case where $|\eta_i| \le \varepsilon |\boldsymbol{a}_i^\top \boldsymbol{x}|^2$,

$$\begin{array}{l} \mathbf{1}_{\left\{|(\boldsymbol{a}_i^\top \boldsymbol{x})^2+\eta_i| \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \le \mathbf{1}_{\left\{(1-\varepsilon)|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \le \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \frac{1+4\varepsilon}{1-\varepsilon}\alpha_y^2\right\}}; \\ \mathbf{1}_{\left\{|(\boldsymbol{a}_i^\top \boldsymbol{x})^2+\eta_i| \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \ge \mathbf{1}_{\left\{(1+\varepsilon)|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \alpha_y^2(\frac{1}{m}\sum_l y_l)\right\}} \ge \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}|^2 \le \frac{1-4\varepsilon}{1+\varepsilon}\alpha_y^2\right\}}. \end{array} \tag{149}$$

Taken collectively, these inequalities imply that

$$\underbrace{\frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i\boldsymbol{a}_i^\top \left(\boldsymbol{a}_i^\top \boldsymbol{x}\right)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \gamma_2\right\}}}_{:=\boldsymbol{Y}_2} \preceq \boldsymbol{Y} \preceq \underbrace{\frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i\boldsymbol{a}_i^\top \left(\boldsymbol{a}_i^\top \boldsymbol{x}\right)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x}| \le \gamma_1\right\}}}_{:=\boldsymbol{Y}_1}. \tag{150}$$

where $\gamma_1 := \max\{\sqrt{(1+4\varepsilon)\alpha_y^2+\varepsilon}, \sqrt{\frac{1+4\varepsilon}{1-\varepsilon}}\alpha_y\}$ and $\gamma_2 := \min\{\sqrt{(1-4\varepsilon)\alpha_y^2-\varepsilon}, \sqrt{\frac{1-4\varepsilon}{1+\varepsilon}}\alpha_y\}$. Letting $\xi \sim \mathcal{N}(0,1)$, one can compute

$$\mathbb{E}[\boldsymbol{Y}_1] = \beta_1 \boldsymbol{x}\boldsymbol{x}^\top + \beta_2 \boldsymbol{I}, \quad \text{and} \quad \mathbb{E}[\boldsymbol{Y}_2] = \beta_3 \boldsymbol{x}\boldsymbol{x}^\top + \beta_4 \boldsymbol{I}, \tag{151}$$

where $\beta_1 := \mathbb{E}\big[\xi^4 \mathbf{1}_{\{|\xi| \le \gamma_1\}}\big] - \mathbb{E}\big[\xi^2 \mathbf{1}_{\{|\xi| \le \gamma_1\}}\big]$, $\beta_2 := \mathbb{E}\big[\xi^2 \mathbf{1}_{\{|\xi| \le \gamma_1\}}\big]$, $\beta_3 := \mathbb{E}\big[\xi^4 \mathbf{1}_{\{|\xi| \le \gamma_2\}}\big] - \mathbb{E}\big[\xi^2 \mathbf{1}_{\{|\xi| \le \gamma_2\}}\big]$ and $\beta_4 := \mathbb{E}\big[\xi^2 \mathbf{1}_{\{|\xi| \le \gamma_2\}}\big]$. Recognizing that $\boldsymbol{a}_i\boldsymbol{a}_i^\top \left(\boldsymbol{a}_i^\top \boldsymbol{x}\right)^2 \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x})| \le c\right\}}$ can be rewritten as $\boldsymbol{b}_i\boldsymbol{b}_i^\top$ for some sub-Gaussian vector $\boldsymbol{b}_i := \boldsymbol{a}_i \left(\boldsymbol{a}_i^\top \boldsymbol{x}\right) \mathbf{1}_{\left\{|\boldsymbol{a}_i^\top \boldsymbol{x})| \le c\right\}}$, we apply standard results on random matrices with non-isotropic sub-Gaussian rows [40, Equation (5.26)] to deduce

$$\|\boldsymbol{Y}_1 - \mathbb{E}[\boldsymbol{Y}_1]\| \le \delta, \quad \|\boldsymbol{Y}_2 - \mathbb{E}[\boldsymbol{Y}_2]\| \le \delta \tag{152}$$

with probability $1 - \exp(-\Omega(m))$, provided that m/n exceeds some large constant. Besides, when ε is sufficiently small, one further has $\|\mathbb{E}[\boldsymbol{Y}_1] - \mathbb{E}[\boldsymbol{Y}_2]\| \le \delta$. These taken collectively with (150) give

$$\|\boldsymbol{Y} - \beta_1 \boldsymbol{x}\boldsymbol{x}^\top - \beta_2 \boldsymbol{I}\| \le 3\delta. \tag{153}$$

Fix some small $\tilde{\delta} > 0$. With (153) in place, applying the Davis-Kahan $\sin\Theta$ theorem [56] and taking δ, ε to be sufficiently small, we obtain

$$\operatorname{dist}(\tilde{\boldsymbol{z}}, \boldsymbol{x}) \le \tilde{\delta},$$

where $\tilde{\boldsymbol{z}}$ is the leading eigenvector of $\boldsymbol{Y}$. Since $\boldsymbol{z}^{(0)} := \lambda_0 \tilde{\boldsymbol{z}}$, one derives

$$\begin{aligned} \operatorname{dist}(\boldsymbol{z}^{(0)}, \boldsymbol{x}) &\le \operatorname{dist}(\lambda_0\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{z}}) + \operatorname{dist}(\tilde{\boldsymbol{z}}, \boldsymbol{x}) \\ &\le |\lambda_0 - 1| + \tilde{\delta} \ \le \ \max\{\sqrt{1+2\varepsilon} - 1,\ 1 - \sqrt{1-2\varepsilon}\} + \tilde{\delta} \end{aligned} \tag{154}$$

as long as m/n is sufficiently large, where the last inequality follows from (147). Picking $\tilde{\delta}$ and ε to be sufficiently small, we establish the claim.

□

We now justify that the Poisson model (4) satisfies the condition (145). Suppose that $\mu_i = (\boldsymbol{a}_i^\top \boldsymbol{x})^2$ and hence $y_i \sim \mathsf{Poisson}(\mu_i)$. It follows from the Chernoff bound that for all $t \ge 0$,

$$\mathbb{P}(y_i - \mu_i \ge \tau) \le \frac{\mathbb{E}[e^{ty_i}]}{\exp(t(\mu_i+\tau))} = \frac{\exp(\mu_i(e^t-1))}{\exp(t(\mu_i+\tau))} = \exp\left(\mu_i\left(e^t - t - 1\right) - t\tau\right),$$

Taking $\tau = 2\tilde{\varepsilon}\mu_i$ and $t = \tilde{\varepsilon}$ for any $0 \le \tilde{\varepsilon} \le 1$ gives

$$\mathbb{P}(y_i - \mu_i \ge 2\tilde{\varepsilon}\mu_i) \ \le \ \exp\left(\mu_i\left(e^t - t - 1 - 2\tilde{\varepsilon}t\right)\right) \overset{\text{(i)}}{\le} \exp\left(\mu_i\left(t^2 - 2\tilde{\varepsilon}t\right)\right) = \exp\left(-\mu_i\tilde{\varepsilon}^2\right),$$

where (i) follows since $e^t \le 1 + t + t^2$ $(0 \le t \le 1)$. In addition, for any $\tilde{\varepsilon} > 1$, taking $t = 1$ we get

$$\mathbb{P}\left(y_i - \mu_i \ge 2\tilde{\varepsilon}\mu_i\right) \;\; \le \;\; \exp\left(\mu_i\left(e^t - t - 1 - 2\tilde{\varepsilon}t\right)\right) \;\; \le \;\; \exp\left(-0.5\mu_i\tilde{\varepsilon}\right).$$

Suppose that $\|\boldsymbol{x}\| \gtrsim \log m$. When $\mu_i \gtrsim \|\boldsymbol{x}\|^2$, setting $\tilde{\varepsilon} = 0.5\varepsilon < 1$ yields

$$\mathbb{P}\left(y_i - \mu_i \ge \varepsilon\mu_i\right) \le \; \exp\left(-\mu_i\varepsilon^2/4\right) \;\; \le \;\; \exp\left(-\Omega(\varepsilon^2\|\boldsymbol{x}\|^2)\right) \;\; = \;\; \exp\left(-\Omega(\varepsilon^2\log^2 m)\right).$$

When $\mu_i \lesssim \|\boldsymbol{x}\|^2$, letting $\kappa_i = \mu_i/\|\boldsymbol{x}\|^2$ and setting $\tilde{\varepsilon} = \varepsilon/2\kappa_i \gtrsim \varepsilon$, we obtain

$$\begin{aligned}\mathbb{P}\left(y_i - \mu_i \ge \varepsilon\|\boldsymbol{x}\|^2\right) &= \mathbb{P}\left(y_i - \mu_i \ge 2\tilde{\varepsilon}\mu_i\right) \le \exp\left(-\min\{\tilde{\varepsilon}, 0.5\}\cdot\tilde{\varepsilon}\mu_i\right) \\ &= \exp\Big(-0.5\min\{\tilde{\varepsilon}, 0.5\}\varepsilon\|\boldsymbol{x}\|^2\Big) = \exp\big(-\Omega\left(\varepsilon^2\log^2 m\right)\big).\end{aligned}$$

In view of the union bound,

$$\mathbb{P}\left(\exists i : \eta_i \ge \varepsilon\max\{\|\boldsymbol{x}\|^2,\; |\boldsymbol{a}_i^\top\boldsymbol{x}|^2\}\right) \;\; \le \;\; m\exp\big(-\Omega\left(\varepsilon^2\log^2 m\right)\big) \;\; \to \;\; 0. \tag{155}$$

Similarly, applying the same argument on $-y_i$ we get $\eta_i \ge -\varepsilon\max\{\|\boldsymbol{x}\|^2, |\boldsymbol{a}_i^\top\boldsymbol{x}|^2\}$ for all i, which together with (155) establishes the condition (145) with high probability. In conclusion, the claim (146) applies to the Poisson model.

D Local error contraction with backtracking line search

In this section, we verify the effectiveness of a backtracking line search strategy by showing local error contraction. To keep it concise, we only sketch the proof for the noiseless case, but the proof extends to the noisy case without much difficulty. Also we do not strive to obtain an optimized constant. For concreteness, we prove the following proposition.

Proposition 4. *The claim in Proposition 1 continues to hold if $\alpha_h \ge 6$, $\alpha_z^{\mathrm{ub}} \ge 5$, $\alpha_z^{\mathrm{lb}} \le 0.1$, $\alpha_p \ge 5$, and*

$$\|\boldsymbol{h}\|/\|\boldsymbol{z}\| \le \epsilon_{\mathrm{tr}} \tag{156}$$

for some constant $\epsilon_{\mathrm{tr}} > 0$ independent of n and m.

Note that if $\alpha_h \ge 6$, $\alpha_z^{\mathrm{ub}} \ge 5$ and $\alpha_z^{\mathrm{lb}} \le 0.1$, then the boundary step size μ_0 given in Proposition 1 satisfies

$$\frac{0.994 - \zeta_1 - \zeta_2 - \sqrt{2/(9\pi)}\alpha_h^{-1}}{2\left(1.02 + 0.665\alpha_h^{-1}\right)} \ge 0.384.$$

Thus, it suffices to show that the step size obtained by a backtracking line search lies within (0,0.384). For notational convenience, we will set

$$\boldsymbol{p} := m^{-1}\nabla\ell_{\mathrm{tr}}\left(\boldsymbol{z}\right) \quad \text{and} \quad \mathcal{E}_3^i := \left\{\left|\boldsymbol{a}_i^\top\boldsymbol{z}\right| \ge \alpha_z^{\mathrm{lb}}\left\|\boldsymbol{z}\right\| \text{ and } \left|\boldsymbol{a}_i^\top\boldsymbol{p}\right| \le \alpha_p\left\|\boldsymbol{p}\right\|\right\}$$

throughout the rest of the proof. We also impose the assumption

$$\left\|\boldsymbol{p}\right\| / \left\|\boldsymbol{z}\right\| \le \epsilon \tag{157}$$

for some sufficiently small constant $\epsilon > 0$, so that $\left|\boldsymbol{a}_i^\top\boldsymbol{p}\right| / \left|\boldsymbol{a}_i^\top\boldsymbol{z}\right|$ is small for all non-truncated terms. It is self-evident from (79) that in the regime under study, one has

$$\left\|\boldsymbol{p}\right\| \ge 2\left\{1.99 - 2\left(\zeta_1 + \zeta_2\right) - \sqrt{8/\pi}(3\alpha_h)^{-1} - o\left(1\right)\right\}\left\|\boldsymbol{h}\right\| \ge 3.64\left\|\boldsymbol{h}\right\|. \tag{158}$$

To start with, consider three scalars h, b, and δ. Setting $b_\delta := \frac{(b+\delta)^2 - b^2}{b^2}$, we get

$$\begin{aligned}&(b+h)^2\log\frac{(b+\delta)^2}{b^2} - (b+\delta)^2 + b^2 = (b+h)^2\log\left(1+b_\delta\right) - b^2 b_\delta \\ &\overset{\text{(i)}}{\le} (b+h)^2\left\{b_\delta - 0.4875b_\delta^2\right\} - b^2 b_\delta \;\; = \;\; \left((b+h)^2 - b^2\right)b_\delta - 0.4875\left(b+h\right)^2 b_\delta^2 \\ &= h\delta\left(2 + h/b\right)\left(2 + \delta/b\right) - 0.4875\left(1 + h/b\right)^2\left|\delta\left(2 + \delta/b\right)\right|^2 \\ &= 4h\delta + \frac{2h^2\delta}{b} + \frac{2h\delta^2}{b} + \frac{h^2\delta^2}{b^2} - 0.4875\delta^2\left(1 + \frac{h}{b}\right)^2\left(2 + \frac{\delta}{b}\right)^2,\end{aligned} \tag{159}$$

where (i) follows from the inequality $\log(1+x) \leq x - 0.4875x^2$ for sufficiently small x. To further simplify the bound, observe that

$$\delta^2 \left(1+\frac{h}{b}\right)^2 \left(2+\frac{\delta}{b}\right)^2 \geq 4\delta^2 \left(1+\frac{h}{b}\right)^2 + \delta^2 \left(1+\frac{h}{b}\right)^2 \frac{4\delta}{b}$$

$$\text{and} \quad \frac{2h\delta^2}{b} + \frac{h^2\delta^2}{b^2} = \left(\left(1+\frac{h}{b}\right)^2 - 1\right)\delta^2.$$

Plugging these two identities into (159) yields

$$\begin{aligned}(159) &\leq 4h\delta + \frac{2h^2\delta}{b} - \left(0.95\left(1+\frac{h}{b}\right)^2 + 1\right)\delta^2 - 0.4875\delta^2\left(1+\frac{h}{b}\right)^2\frac{4\delta}{b} \\ &\leq 4h\delta - 1.95\delta^2 + \frac{2h^2|\delta|}{|b|} + \frac{1.9|h|}{|b|}\delta^2 + \frac{1.95\left|\delta^3\right|}{|b|}\left(1+\frac{h}{b}\right)^2.\end{aligned}$$

Replacing respectively b, δ, and h with $\boldsymbol{a}_i^\top \boldsymbol{z}$, $\tau \boldsymbol{a}_i^\top \boldsymbol{p}$, and $-\boldsymbol{a}_i^\top \boldsymbol{h}$, one sees that the log-likelihood $\ell_i(\boldsymbol{z}) = y_i \log(|\boldsymbol{a}_i^\top \boldsymbol{z}|^2) - |\boldsymbol{a}_i^\top \boldsymbol{z}|^2$ obeys

$$\begin{aligned}\ell_i(\boldsymbol{z}+\tau\boldsymbol{p}) - \ell_i(\boldsymbol{z}) &= y_i \log\frac{\left|\boldsymbol{a}_i^\top(\boldsymbol{z}+\tau\boldsymbol{p})\right|^2}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2} - \left|\boldsymbol{a}_i^\top(\boldsymbol{z}+\tau\boldsymbol{p})\right|^2 + \left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|^2 \\ &\leq \underbrace{-4\tau\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)\left(\boldsymbol{a}_i^\top \boldsymbol{p}\right)}_{:=I_{1,i}} \underbrace{- 1.95\tau^2\left(\boldsymbol{a}_i^\top \boldsymbol{p}\right)^2}_{:=I_{2,i}} + \underbrace{\frac{2\tau\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2\left|\boldsymbol{a}_i^\top \boldsymbol{p}\right|}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}}_{:=I_{3,i}} + \underbrace{\frac{1.9\tau^2\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\left(\boldsymbol{a}_i^\top \boldsymbol{p}\right)^2}_{:=I_{4,i}} \\ &\quad + \underbrace{\frac{1.95\tau^3\left|\boldsymbol{a}_i^\top \boldsymbol{p}\right|^3}{\left|\boldsymbol{a}_i^\top \boldsymbol{z}\right|}\left(1-\frac{\boldsymbol{a}_i^\top \boldsymbol{h}}{\boldsymbol{a}_i^\top \boldsymbol{z}}\right)^2}_{:=I_{5,i}}.\end{aligned}$$

The next step is then to bound each of these terms separately. Most of the following bounds are straightforward consequences from [6, Lemma 3.1] combined with the truncation rule. For the first term, applying the AM-GM inequality we get

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m I_{1,i}\mathbf{1}_{\mathcal{E}_3^i} &\leq \frac{4\tau}{3.64m}\sum_{i=1}^m\left\{\frac{3.64^2}{2}\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2 + \frac{1}{2}\left(\boldsymbol{a}_i^\top \boldsymbol{p}\right)^2\right\} \\ &\leq \frac{4\tau(1+\delta)}{3.64}\left\{\frac{3.64^2}{2}\|\boldsymbol{h}\|^2 + \frac{1}{2}\|\boldsymbol{p}\|^2\right\}.\end{aligned}$$

Secondly, it follows from Lemma 4 that

$$\frac{1}{m}\sum_{i=1}^m I_{2,i}\mathbf{1}_{\mathcal{E}_3^i} = -1.95\tau^2\frac{1}{m}\sum_{i=1}^m\left(\boldsymbol{a}_i^\top \boldsymbol{p}\right)^2\mathbf{1}_{\mathcal{E}_3^i} \leq -1.95\left(1-\tilde{\zeta}_1-\tilde{\zeta}_2\right)\tau^2\|\boldsymbol{p}\|^2,$$

where $\tilde{\zeta}_1 := \max\{\mathbb{E}\big[\xi^2\mathbf{1}_{\left\{|\xi|\leq\sqrt{1.01}\alpha_z^{\mathrm{lb}}\right\}}\big], \mathbb{E}\big[\mathbf{1}_{\left\{|\xi|\leq\sqrt{1.01}\alpha_z^{\mathrm{lb}}\right\}}\big]\}$ and $\tilde{\zeta}_2 := \mathbb{E}\big[\xi^2\mathbf{1}_{\left\{|\xi|>\sqrt{0.99}\alpha_h\right\}}\big]$. The third term is controlled by

$$\frac{1}{m}\sum_{i=1}^m I_{3,i}\mathbf{1}_{\mathcal{E}_3^i} \leq 2\tau\frac{\alpha_p\|\boldsymbol{p}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}\left\{\frac{1}{m}\sum_{i=1}^m\left(\boldsymbol{a}_i^\top \boldsymbol{h}\right)^2\right\} \lesssim \tau\epsilon\|\boldsymbol{h}\|^2.$$

Fourthly, it arises from the AM-GM inequality that

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^m I_{4,i}\mathbf{1}_{\mathcal{E}_3^i} &\leq \frac{1.9\tau^2\alpha_p\|\boldsymbol{p}\|}{\alpha_z^{\mathrm{lb}}\|\boldsymbol{z}\|}\frac{1}{m}\sum_{i=1}^m\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|\left|\boldsymbol{a}_i^\top \boldsymbol{p}\right| \lesssim \epsilon\tau^2\frac{1}{m}\sum_{i=1}^m\left\{2\left|\boldsymbol{a}_i^\top \boldsymbol{h}\right|^2 + \frac{1}{8}\left|\boldsymbol{a}_i^\top \boldsymbol{p}\right|^2\right\} \\ &\lesssim \epsilon\tau^2\|\boldsymbol{p}\|^2.\end{aligned}$$

Finally, the last term is bounded by

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m} I_{5,i}\mathbf{1}_{\mathcal{E}_3^i} &\leq \frac{1}{m}\sum_{i=1}^{m}\frac{1.95\tau^3\left|\boldsymbol{a}_i^\top\boldsymbol{p}\right|^3}{\left|\boldsymbol{a}_i^\top\boldsymbol{z}\right|}\left(\frac{\boldsymbol{a}_i^\top\boldsymbol{x}}{\boldsymbol{a}_i^\top\boldsymbol{z}}\right)^2 \leq \frac{1.95\tau^3\alpha_p^3\left\|\boldsymbol{p}\right\|^3}{(\alpha_z^{\mathrm{lb}})^3\left\|\boldsymbol{z}\right\|^3}\frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{a}_i^\top\boldsymbol{x}\right)^2 \\ &\lesssim \tau^3\epsilon\frac{\left\|\boldsymbol{x}\right\|^2}{\left\|\boldsymbol{z}\right\|^2}\left\|\boldsymbol{p}\right\|^2.\end{aligned}$$

Under the hypothesis (158), we can further derive $\frac{1}{m}\sum_{i=1}^{m} I_{1,i}\mathbf{1}_{\mathcal{E}_3^i} \leq \tau\left(1.1+\delta\right)\left\|\boldsymbol{p}\right\|^2$. Putting all the above bounds together yields that the truncated objective function is majorized by

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\left\{\ell_i\left(\boldsymbol{z}+\tau\boldsymbol{p}\right)-\ell_i\left(\boldsymbol{z}\right)\right\}\mathbf{1}_{\mathcal{E}_3^i} \leq \frac{1}{m}\sum_{i=1}^{m}\left(I_{1,i}+I_{2,i}+I_{3,i}+I_{4,i}+I_{5,i}\right)\mathbf{1}_{\mathcal{E}_3^i} \\ &\leq \tau\left(1.1+\delta\right)\left\|\boldsymbol{p}\right\|^2-1.95\left(1-\tilde{\zeta}_1-\tilde{\zeta}_2\right)\tau^2\left\|\boldsymbol{p}\right\|^2+\tau\tilde{\epsilon}\left\|\boldsymbol{p}\right\|^2 \\ &= \left\{\tau\left(1.1+\delta\right)-1.95\left(1-\tilde{\zeta}_1-\tilde{\zeta}_2\right)\tau^2+\tau\tilde{\epsilon}\right\}\left\|\boldsymbol{p}\right\|^2 \qquad (160)\end{aligned}$$

for some constant $\tilde{\epsilon}>0$ that is linear in ϵ.

Note that the backtracking line search seeks a point satisfying $\frac{1}{m}\sum_{i=1}^{m}\left\{\ell_i\left(\boldsymbol{z}+\tau\boldsymbol{p}\right)-\ell_i\left(\boldsymbol{z}\right)\right\}\mathbf{1}_{\mathcal{E}_3^i} \geq \frac{1}{2}\tau\left\|\boldsymbol{p}\right\|^2$. Given the above majorization (160), this search criterion is satisfied only if

$$\tau/2 \leq \tau\left(1.1+\delta\right)-1.95(1-\tilde{\zeta}_1-\tilde{\zeta}_2)\tau^2+\tau\tilde{\epsilon}$$

or, equivalently,

$$\tau \leq \frac{0.6+\delta+\tilde{\epsilon}}{1.95(1-\tilde{\zeta}_1-\tilde{\zeta}_2)} := \tau_{\mathrm{ub}}.$$

Taking δ and $\tilde{\epsilon}$ to be sufficiently small, we see that $\tau \leq \tau_{\mathrm{ub}} \leq 0.384$, provided that $\alpha_z^{\mathrm{lb}} \leq 0.1$, $\alpha_z^{\mathrm{ub}} \geq 5$, $\alpha_h \geq 6$, and $\alpha_p \geq 5$.

Using very similar arguments, one can also show that $\frac{1}{m}\sum_{i=1}^{m}\left\{\ell_i\left(\boldsymbol{z}+\tau\boldsymbol{p}\right)-\ell_i\left(\boldsymbol{z}\right)\right\}\mathbf{1}_{\mathcal{E}_3^i}$ is minorized by a similar quadratic function, which combined with the stopping criterion $\frac{1}{m}\sum_{i=1}^{m}\left\{\ell_i\left(\boldsymbol{z}+\tau\boldsymbol{p}\right)-\ell_i\left(\boldsymbol{z}\right)\right\}\mathbf{1}_{\mathcal{E}_3^i} \geq \frac{1}{2}\tau\left\|\boldsymbol{p}\right\|^2$ suggests that τ is bounded away from 0. We omit this part for conciseness.

Acknowledgements

E. C. is partially supported by NSF under grant CCF-0963835 and by the Math + X Award from the Simons Foundation. Y. C. is supported by the same NSF grant. We thank Carlos Sing-Long and Weijie Su for their constructive comments about an early version of the manuscript. E. C. is grateful to Xiaodong Li and Mahdi Soltanolkotabi for many discussions about Wirtinger flows. We thank the anonymous reviewers for helpful comments.

References

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

[2] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization*, volume 2. SIAM, 2001.

[3] R. W. Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237, 1972.

[4] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21:2758–2769, 1982.

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

[6] E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1017–1026, 2013.

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

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

[9] Y. Chen, Y. Chi, and A. J. Goldsmith. Exact and stable covariance estimation from quadratic sampling via convex programming. *IEEE Transactions on Information Theory*, 61(7):4034–4059, 2015.

[10] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev. Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing. *Optics express*, 19(16), 2011.

[11] H. Ohlsson, A. Yang, R. Dong, and S. S. Sastry. Compressive phase retrieval from squared output measurements via semidefinite programming. *Neural Information Processing Systems (NIPS)*, 2011.

[12] L. Demanet and P. Hand. Stable optimizationless recovery from phaseless linear measurements. *Journal of Fourier Analysis and Applications*, 20(1):199–221, 2014.

[13] X. Li and V. Voroninski. Sparse signal recovery from quadratic measurements via convex programming. *SIAM Journal on Mathematical Analysis*, 45(5):3019–3033, 2013.

[14] T. Cai and A. Zhang. ROP: Matrix recovery via rank-one projections. *The Annals of Statistics*, 43(1):102–138, 2015.

[15] K. Jaganathan, S. Oymak, and B. Hassibi. Recovery of sparse 1-D signals from the magnitudes of their Fourier transform. In *IEEE ISIT*, pages 1473–1477, 2012.

[16] Y. Chen, X. Yi, and C. Caramanis. A convex formulation for mixed regression with two components: Minimax optimal rates. In *Conf. on Learning Theory*, 2014.

[17] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements. *arXiv preprint arXiv:1410.6913*, 2014.

[18] S. Bahmani and J. Romberg. Efficient compressive phase retrieval with constrained sensing vectors. *Advances in Neural Information Processing Systems*, pages 523–531, 2015.

[19] D. Gross, F. Krahmer, and R. Kueng. A partial derandomization of phaselift using spherical designs. *Journal of Fourier Analysis and Applications*, 21(2):229–266, 2015.

[20] D. Gross, F. Krahmer, and R. Kueng. Improved recovery guarantees for phase retrieval from coded diffraction patterns. *arXiv preprint arXiv:1402.6286*, 2014.

[21] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[22] V. Elser. Phase retrieval by iterated projections. *JOSA. A*, 20(1):40–55, 2003.

[23] P. Schniter and S. Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, Feb 2015.

[24] Q. Zheng and J. Lafferty. A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. *Advances in Neural Information Processing Systems*, 2015.

[25] S. Marchesin, Y. Tu, and H. Wu. Alternating projection, ptychographic imaging and phase synchronization. *arXiv:1402.0550*, 2014.

[26] A. Repetti, E. Chouzenoux, and J-C Pesquet. A nonconvex regularized approach for phase retrieval. *International Conference on Image Processing*, pages 1753–1757, 2014.

[27] Y. Shechtman, A. Beck, and Y. C. Eldar. GESPAR: Efficient phase retrieval of sparse signals. *IEEE Transactions on Signal Processing*, 62(4):928–938, 2014.

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

[29] M. Soltanolkotabi. *Algorithms and Theory for Clustering and Nonconvex Quadratic Programming*. PhD thesis, Stanford University, 2014.

[30] E. J. Candès, X. Li, and M. Soltanolkotabi. `http://www-bcf.usc.edu/~soltanol/PRWF.html`, 2014.

[31] J. Nocedal and S. J. Wright. *Numerical Optimization (2nd edition)*. Springer, 2006.

[32] L. N. Trefethen and D. Bau III. *Numerical linear algebra*, volume 50. SIAM, 1997.

[33] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *to appear in Applied and Computational Harmonic Analysis*, 2014.

[34] Y. Chen and E. J. Candès. Supplemental materials for: "solving random quadratic systems of equations is nearly as easy as solving linear systems". May 2015.

[35] K. Wei. Phase retrieval via Kaczmarz methods. *arXiv preprint arXiv:1502.01822*, 2015.

[36] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. Phase retrieval with polarization. *SIAM Journal on Imaging Sciences*, 7(1):35–66, 2014.

[37] R. Balan, B. Bodmann, P. Casazza, and D. Edidin. Painless reconstruction from magnitudes of frame coefficients. *Journal of Fourier Analysis and Applications*, 15(4):488–501, 2009.

[38] Y. C. Eldar and S. Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*, 36(3):473–494, 2014.

[39] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson. Saving phase: Injectivity and stability for phase retrieval. *Applied and Computational Harmonic Analysis*, 37(1):106–125, 2014.

[40] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing, Theory and Applications*, pages 210 – 268, 2012.

[41] A. B. Tsybakov and V. Zaiats. *Introduction to nonparametric estimation*, volume 11. Springer, 2009.

[42] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980 –2998, June 2010.

[43] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *ACM symposium on Theory of computing*, pages 665–674, 2013.

[44] M. Hardt and M. Wootters. Fast matrix completion without the condition number. *Conference on Learning Theory*, pages 638 – 678, 2014.

[45] R. Sun and Z. Luo. Guaranteed matrix completion via non-convex factorization. *arXiv:1411.8003*, 2014.

[46] P. Netrapalli, U. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain. Non-convex robust PCA. In *Advances in Neural Information Processing Systems*, pages 1107–1115, 2014.

[47] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery using nonconvex optimization. *International Conference on Machine Learning*, pages 2351–2360, 2015.

[48] J. Sun, Q. Qu, and J. Wright. When are nonconvex problems not scary? *arXiv preprint arXiv:1510.06096*, 2015.

[49] S. Arora, R. Ge, T. Ma, and A. Moitra. Simple, efficient, and neural algorithms for sparse coding. *arXiv:1503.00778*, 2015.

[50] X. Yi, C. Caramanis, and S. Sanghavi. Alternating minimization for mixed linear regression. *International Conference on Machine Learning*, June 2014.

[51] S. Balakrishnan, M. J. Wainwright, and B. Yu. Statistical guarantees for the EM algorithm: From population to sample-based analysis. *arXiv preprint arXiv:1408.2156*, 2014.

[52] A. S. Bandeira, M. Charikar, A. Singer, and A. Zhu. Multireference alignment using semidefinite programming. In *Conference on Innovations in theoretical computer science*, pages 459–470, 2014.

[53] Q. Huang and L. Guibas. Consistent shape maps via semidefinite programming. *Computer Graphics Forum*, 32(5):177–186, 2013.

[54] Y. Chen, L. J. Guibas, and Q. Huang. Near-optimal joint optimal matching via convex relaxation. *International Conference on Machine Learning (ICML)*, pages 100 – 108, June 2014.

[55] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, 2005.

[56] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.