Ann. Appl. Math.
doi: 10.4208/aam.OA-2021-0009

Vol. **37**, No. 4, pp. 437-512
November 2021

The Global Landscape of Phase Retrieval I: Perturbed Amplitude Models

Jian-Feng Cai[1], Meng Huang[1], Dong Li[2,*] and Yang Wang[1]

[1] *Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

[2] *SUSTech International Center for Mathematics and Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China*

Received 3 April 2021; Accepted (in revised version) 10 November 2021

Abstract. A fundamental task in phase retrieval is to recover an unknown signal $\boldsymbol{x}\in\mathbb{R}^n$ from a set of magnitude-only measurements $y_i=|\langle \boldsymbol{a}_i,\boldsymbol{x}\rangle|$, $i=1,\cdots,m$. In this paper, we propose two novel perturbed amplitude models (PAMs) which have a non-convex and quadratic-type loss function. When the measurements $\boldsymbol{a}_i\in\mathbb{R}^n$ are Gaussian random vectors and the number of measurements $m\geq Cn$, we rigorously prove that the PAMs admit no spurious local minimizers with high probability, i.e., the target solution $\boldsymbol{x}$ is the unique local minimizer (up to a global phase) and the loss function has a negative directional curvature around each saddle point. Thanks to the well-tamed benign geometric landscape, one can employ the vanilla gradient descent method to locate the global minimizer $\boldsymbol{x}$ (up to a global phase) without spectral initialization. We carry out extensive numerical experiments to show that the gradient descent algorithm with random initialization outperforms state-of-the-art algorithms with spectral initialization in empirical success rate and convergence speed.

AMS subject classifications: 94A12, 65K10, 49K45

Key words: Phase retrieval, landscape analysis, non-convex optimization.

---

1 Introduction

**Emails:* `jfcai@ust.hk` (J. Cai), `menghuang@ust.hk` (M. Huang), `lid@sustech.edu.cn` (D. Li), `yangwang@ust.hk` (Y. Wang)

1.1 Background

The basic amplitude model for phase retrieval can be written as

$$y_j = |\langle \boldsymbol{a}_j, \boldsymbol{x}\rangle|, \quad j=1,\cdots,m,$$

where $\boldsymbol{a}_j \in \mathbb{R}^n$, $j=1,\cdots,m$ are given vectors and m is the number of measurements. The goal is to recover the unknown signal $\boldsymbol{x}\in\mathbb{R}^n$ based on the measurements $\{(\boldsymbol{a}_j,y_j)\}_{j=1}^m$. This problem arises in many fields of science and engineering such as X-ray crystallography [16, 24], microscopy [23], astronomy [7], coherent diffractive imaging [15,28] and optics [34] etc. In practical applications due to the physical limitations optical detectors can only record the magnitude of signals while losing the phase information. Despite its simple mathematical formulation, it has been shown that reconstructing a finite-dimensional discrete signal from the magnitude of its Fourier transform is generally an *NP-complete* problem [27].

Many algorithms have been designed to solve the phase retrieval problem, which can be categorized into convex algorithms and non-convex ones. The convex algorithms usually rely on a "matrix-lifting" technique, which lifts the phase retrieval problem into a low rank matrix recovery problem. By using convex relaxation one can recast the matrix recovery problem as a convex optimization problem. The corresponding algorithms include PhaseLift [2, 4], PhaseCut [33] etc. It has been shown [2] that PhaseLift can achieve the exact recovery under the optimal sampling complexity with Gaussian random measurements.

Although convex methods have good theoretical guarantees of convergence, they tend to be computationally inefficient for large scale problems. In contrast, many non-convex algorithms bypass the lifting step and operate directly on the lower-dimensional ambient space, making them much more computationally efficient. Early non-convex algorithms were mostly based on the technique of alternating projections, e.g., Gerchberg-Saxton [14] and Fineup [9]. The main drawback, however, is the lack of theoretical guarantee. Later Netrapalli et al. [25] proposed the AltMinPhase algorithm based on a technique known as *spectral initialization*. They proved that the algorithm linearly converges to the true solution with $\mathcal{O}(n\log^3 n)$ resampling Gaussian random measurements. This work led further to several other non-convex algorithms based on spectral initialization. A common thread is first choosing a good initial guess through spectral initialization, and then solving an optimization model through gradient descent. Two widely used optimization estimators are the intensity-based loss

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} F(\boldsymbol{z}) = \sum_{j=1}^m \left(|\langle \boldsymbol{a}_j,\boldsymbol{z}\rangle|^2 - y_j^2\right)^2; \tag{1.1}$$

and the amplitude-based loss

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} F(\boldsymbol{z})=\sum_{j=1}^{m}(|\langle \boldsymbol{a}_j,\boldsymbol{z}\rangle|-y_j)^2. \tag{1.2}$$

Specifically, Candès et al. developed the Wirtinger Flow (WF) method [3] based on (1.1) and proved that the WF algorithm can achieve linear convergence with $\mathcal{O}(n\log n)$ Gaussian random measurements. Chen and Candès in [6] improved the results to $\mathcal{O}(n)$ Gaussian random measurements by incorporating a truncation which leads to a novel Truncated Wirtinger Flow (TWF) algorithm. Other methods based on (1.1) include the Gauss-Newton method [11], the trust-region method [29] and the like [17]. For the amplitude flow estimator (1.2), several algorithms have also been developed recently, such as the Truncated Amplitude Flow (TAF) algorithm [35], the Reshaped Wirtinger Flow (RWF) [37] algorithm, randomized Kaczmarz methods [18, 19, 31, 36] and the Perturbed Amplitude Flow (PAF) [10] algorithm. Those algorithms have been shown to converge linearly to the true solution up to a global phase with $\mathcal{O}(n)$ Gaussian random measurements. Furthermore, there is ample evidence from numerical simulations showing that algorithms based on the amplitude flow loss (1.2) tend to outperform algorithms based on loss (1.1) when measured in empirical success rate and convergence speed.

1.2 Prior arts and connections

As was already mentioned earlier, producing a good initial guess using spectral initialization seems to be a prerequisite for prototypical non-convex algorithms to succeed with good theoretical guarantees. A natural and fundamental question is:

Is it possible for non-convex algorithms to achieve successful recovery with a random initialization (*i.e., without spectral initialization or any additional truncation*)?

For the intensity-based estimator (1.1), the answer is affirmative. In the recent work [29], Ju Sun et al.. carried out a deep study of the global geometric structure of the loss function of (1.1). They proved that the loss function $F(\boldsymbol{z})$ does not have any spurious local minima under $\mathcal{O}(n\log^3 n)$ Gaussian random measurements. More specifically, it was shown in [29] that all minimizers coincide with the target signal $\boldsymbol{x}$ up to a global phase, and the loss function has a negative directional curvature around each saddle point. Thanks to this benign geometric landscape any algorithm which can avoid saddle points converges to the true solution with high probability. A trust-region method was employed in [29] to find the global minimizers with random initialization. To reduce the sampling complexity, it has been shown in [22] that a combination of the loss function (1.1) with a judiciously chosen activation function also possesses the benign geometry structure under $\mathcal{O}(n)$ Gaussian random measurements. Recently, a smoothed amplitude flow estimator has been proposed in [5] and the authors show that the loss function has benign geometry structure under the optimal sampling complexity. Numerical tests show that the estimator in [5] yields very stable and fast convergence with random initialization and performs as good as or even better than the existing gradient descent methods with spectral initialization.

The emerging concept of a benign geometric landscape has also recently been explored in many other applications of signal processing and machine learning, e.g., matrix sensing [1,26], tensor decomposition [12], dictionary learning [30] and matrix completion [13]. For general optimization problems there exist a plethora of loss functions with well-behaved geometric landscapes such that all local optima are also global optima and each saddle point has a negative direction curvature in its vincinity. Correspondingly several techniques have been developed to guarantee that the standard gradient based optimization algorithms can escape such saddle points efficiently, see e.g., [8,20,21].

1.3 Our contributions

This paper aims to give a positive answer to the problem proposed in Subsection 1.2, especially for the amplitude-based model. We first introduce two novel estimators based on a deep modification of (1.2) and then we prove rigorously that their loss functions have a benign geometric landscape under the optimal sampling complexity $\mathcal{O}(n)$, namely, the loss functions have no spurious local minimizers and have a negative directional curvature around each saddle point. Such properties allow first order method like gradient descent to find a global minimum with random initial guess. We carry out extensive numerical experiments and show that the gradient descent algorithm with random initialization outperforms the state-of-the-art algorithms with spectral initialization in empirical success rate and convergence speed.

We now give a slightly more detailed summary of the main theoretical results proved in our papers. Consider the loss function which is akin to the estimator (1.2):

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} f(\boldsymbol{z})=\frac{1}{m}\sum_{i=1}^{m}\left(\sqrt{\beta|\boldsymbol{z}|^2+(\boldsymbol{a}_i^T\boldsymbol{z})^2}-\sqrt{\beta|\boldsymbol{z}|^2+y_i^2}\right)^2. \tag{1.3}$$

The following theorem shows that the loss function above has benign geometry structure under optimal sampling complexity.

Theorem 1.1 (Informal)**.** *Consider the perturbed amplitude model (PAM1)* (1.3)*. Assume* $\{\boldsymbol{a}_i\}_{i=1}^m$ *are i.i.d. Gaussian random vectors and* $\boldsymbol{x}\neq 0$*. Let* $0<\beta<\infty$*. If*

$m\geq Cn$, then with probability at least $1-\mathcal{O}(m^{-2})$, the loss function $f=f(\boldsymbol{z})$ has no spurious local minimizers. The only local minimizer is $\pm\boldsymbol{x}$, and all saddle points are strict saddles.

The avid reader should notice that the probability concentration in Theorem 1.1 is only $1-\mathcal{O}(m^{-2})$. Besides, the function is only Lipschitz continuous near the origin[†]. To remedy this and improve the probability of success, we introduce the following genuinely globally smooth estimator:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} f(\boldsymbol{z})=\frac{1}{m}\sum_{i=1}^{m}\left(\sqrt{\beta|\boldsymbol{z}|^2+(\boldsymbol{a}_i^T\boldsymbol{z})^2+y_i^2}-\sqrt{\beta|\boldsymbol{z}|^2+2y_i^2}\right)^2. \tag{1.4}$$

The geometric landscape is stated below.

Theorem 1.2 (Informal)**.** *Consider the perturbed amplitude model (PAM2)* (1.4). *Assume $\{\boldsymbol{a}_i\}_{i=1}^m$ are i.i.d. Gaussian random vectors and $\boldsymbol{x}\neq 0$. Let $0<\beta<\infty$. If $m\geq Cn$, then with probability at least $1-e^{-cm}$, the loss function $f=f(\boldsymbol{z})$ has no spurious local minimizers. The only local minimizer is $\pm\boldsymbol{x}$, and all other critical points are strict saddles.*

Remark 1.1. In a con-current work [5], we considered another new smoothed amplitude-based estimator which is based on a piece-wise smooth modification of the amplitude estimator (1.2). The estimator takes the form

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} F(\boldsymbol{z})=\frac{1}{2m}\sum_{i=1}^{m}y_i^2\left(\gamma\left(\frac{\boldsymbol{a}_i^\top\boldsymbol{z}}{y_i}\right)-1\right)^2,$$

where the function $\gamma(t)$ is taken to be

$$\gamma(t):=\begin{cases}|t|, & |t|>\beta,\\ \dfrac{1}{2\beta}t^2+\dfrac{\beta}{2}, & |t|\leq\beta.\end{cases}$$

For $0<\beta\leq 1/2$, we prove that the loss function has a benign landscape under the optimal sampling threshold $m=\mathcal{O}(n)$. There are subtle technical difficulties in connection with the piecewise-smoothness of the loss function which make the overall proof therein quite special. On the other hand, there are exciting evidences that the machinery developed in this work can be generalized significantly in various directions (including complex-valued cases etc.). We plan to address some of these important issues in forthcoming works.

[†]This is due to the function $\sqrt{\beta|\boldsymbol{z}|^2+(\boldsymbol{a}_i^\top\boldsymbol{z})^2}$.

1.4 Organization

The paper is organized as follows. In Section 2, we analyze the global geometric structure for the first estimator, and the global analysis for the second estimator is given in Section 3. For both estimators, we show that their loss functions have no spurious local minimizers under optimal sampling complexity $\mathcal{O}(n)$. In Section 4, we give some numerical experiments to demonstrate the efficiency of our proposed estimators. In Appendix, we collect the technique lemmas which are used in the proof.

1.5 Notations

Throughout this proof we fix $\beta>0$ as a constant and do not study the precise dependence of other parameters on β. We write $u\in\mathbb{S}^{n-1}$ if $u\in\mathbb{R}^n$ and $\|u\|_2=\sqrt{\sum_j(u_j)^2}=1$. We use χ to denote the usual characteristic function. For example $\chi_A(x)=1$ if $x\in A$ and $\chi_A(x)=0$ if $x\notin A$. We denote by δ_1, ϵ, η, η_1 various constants whose value will be taken sufficiently small. The needed smallness will be clear from the context. For any quantity X, we shall write

$$X=\mathcal{O}(Y) \quad \text{if} \quad |X|\le CY$$

for some constant $C>0$. We write $X\lesssim Y$ if $X\le CY$ for some constant $C>0$. We shall write $X\ll Y$ if $X\le cY$ where the constant $c>0$ will be sufficiently small. In our proof it is important for us to specify the precise dependence of the sampling size m in terms of the dimension n. For this purpose we shall write $m\gtrsim n$ if $m\ge Cn$ where the constant C is allowed to depend on β and the small constants ϵ, ϵ_i etc used in the argument. One can extract more explicit dependence of C on the small constants and β but for simplicity we suppress this dependence here. We shall say an event A happens with **high probability** if

$$\mathbb{P}(A)\ge 1-Ce^{-cm},$$

where $c>0$, $C>0$ are constants. The constants c and C are allowed to depend on β and the small constants ϵ, δ mentioned before.

2 Perturbed amplitude model I

Recall the loss function of perturbed amplitude model (PAM1) (1.3):

$$f(u)=\frac{1}{m}\sum_{j=1}^m\Big(\sqrt{\beta|u|^2+(a_j\cdot u)^2}-\sqrt{\beta|u|^2+(a_j\cdot x)^2}\,\Big)^2, \tag{2.1}$$

where $\beta>0$ is a parameter. Here, we denote

$$|u|:=\|u\|_2=\sqrt{\sum_j u_j^2}$$

for the convenience and write $\boldsymbol{a}_i$ as a_i, $\boldsymbol{x}$ as x to alleviate the notation. The global geometric structure of above empirical loss is stated below.

Theorem 2.1. *Let $0<\beta<\infty$. Assume $\{a_i\}_{i=1}^m$ are i.i.d. standard Gaussian random vectors and $x\neq 0$. There exist positive constants C, C_1 depending only on β, such that if $m\geq Cn$, then with probability at least $1-\frac{C_1}{m^2}$ the loss function $f(u)$ defined by* (2.1) *has no spurious local minimizers. The only local minimizer is $\pm x$, and the loss function is strongly convex in a neighborhood of $\pm x$. At the point $u=0$ the loss function has non-vanishing directional gradient along any direction $\xi\in\mathbb{S}^{n-1}$. All other critical points are strict saddles, i.e., each saddle point has a neighborhood where the function has negative directional curvature.*

Remark 2.1. We shall show that most of the statements can be proved with high probability $1-e^{-cm}$. The only part where the weaker probability $1-\mathcal{O}(m^{-2})$ is used comes in the analysis of the strong convexity near the global minimizer $u=\pm x$ (see e.g., Lemma A.10). This can be refined but we shall not dwell on it here.

In view of this homogeneity and the rotation invariance of the Gaussian distribution, we may assume without loss of generality that $x=e_1$ when studying the landscape of $f(u)$. Thus throughout the rest of the proof we shall assume $x=e_1$.

2.1 The regimes $\|u\|_2\leq\frac{\sqrt{\beta}}{4(1+\beta)}$ and $\|u\|_2\geq 3\sqrt{1+\beta}$ are fine

Write $u=\rho\hat{u}$, where $\hat{u}\in\mathbb{S}^{n-1}$. Then

$$\begin{aligned} f_j(u,e_1)=&\Big(\sqrt{\beta|u|^2+(a_j\cdot u)^2}-\sqrt{\beta|u|^2+(a_j\cdot e_1)^2}\Big)^2\\ =&\rho^2((a_j\cdot\hat{u})^2+2\beta)+(a_j\cdot e_1)^2-2\rho\sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}. \end{aligned}$$

The derivative with respect to ρ is

$$\begin{aligned} \partial_\rho f=\frac{1}{m}\sum_{j=1}^m\Big(&2\rho((a_j\cdot\hat{u})^2+2\beta)-2\sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}\\ &-\frac{2\beta\rho^2\sqrt{\beta+(a_j\cdot\hat{u})^2}}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}}\Big). \end{aligned} \tag{2.2}$$

Lemma 2.1 (The regime $\rho\geq 3\sqrt{1+\beta}$ is OK)**.** *For $m\gtrsim n$, with high probability it holds that*

$$\partial_\rho f>0,\quad \forall \rho\geq 3\sqrt{1+\beta},\quad \forall \hat{u}\in\mathbb{S}^{n-1}.$$

Proof. To prove this lemma, we need to lower bound the first term and upper bound the last two terms of $\partial_\rho f$. For the first term, by using Bernstein's inequality, we have with high probability,

$$\Big|\frac{1}{m}\sum_{j=1}^m (a_j\cdot\hat{u})^2-1\Big|\leq\delta_1\ll 1,\quad \forall\hat{u}\in\mathbb{S}^{n-1}. \tag{2.3}$$

It immediately gives

$$\frac{1}{m}\sum_{j=1}^m\Big(2\rho((a_j\cdot\hat{u})^2+2\beta)\Big)\geq 2\rho((1-\delta_1)+2\beta).$$

For the second term, simple calculation leads to

$$\begin{aligned}&\frac{2}{m}\sum_{j=1}^m\Big(\sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}\Big)\\ \leq&\frac{2}{m}\Big(\sum_{j=1}^m(\beta+(a_j\cdot\hat{u})^2)\Big)^{\frac12}\cdot\Big(\sum_{j=1}^m(\beta\rho^2+(a_j\cdot e_1)^2)\Big)^{\frac12}\\ \leq&2(\beta+1+\delta_1)^{\frac12}(\beta\rho^2+1+\delta_1)^{\frac12}.\end{aligned}$$

Finally, it is easy to derive from (2.3) that

$$\begin{aligned}&\frac{2}{m}\sum_{j=1}^m\Big(\sqrt{\beta+(a_j\cdot\hat{u})^2}\frac{\beta\rho^2}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}}\Big)\\ \leq&\frac{2}{m}\sum_{j=1}^m\Big(\sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}\Big)\\ \leq&2(\beta+1+\delta_1)^{\frac12}(\beta\rho^2+1+\delta_1)^{\frac12}.\end{aligned}$$

Putting all above estimators into (2.2) gives

$$\begin{aligned}\partial_\rho f\geq&2\rho((1-\delta_1)+2\beta)-4(\beta+1+\delta_1)^{\frac12}(\beta\rho^2+1+\delta_1)^{\frac12}\\ =&2\cdot\frac{\rho^2(1-(2+8\beta)\delta_1)-4(1+\beta)+(-8-4\beta)\delta_1+(\rho^2-4)\delta_1^2}{\rho((1-\delta_1)+2\beta)+2(\beta+1+\delta_1)^{\frac12}(\beta\rho^2+1+\delta_1)^{\frac12}}.\end{aligned}$$

Clearly if $\delta_1>0$ is sufficiently small and $\rho\geq 3\sqrt{1+\beta}$, then $\partial_\rho f>0$. $\square$

Lemma 2.2 (The regime $0<\rho\leq\frac{\sqrt{\beta}}{4(1+\beta)}$ is OK)**.** *For $m\gtrsim n$, with high probability it holds that*

$$\partial_\rho f<0,\quad \forall 0<\rho\leq\frac{\sqrt{\beta}}{4(1+\beta)},\quad \forall\hat{u}\in\mathbb{S}^{n-1}.$$

Proof. By Bernstein's inequality, we have with high probability,

$$\begin{aligned}&\Big|\frac{1}{m}\sum_{j=1}^m|a_j\cdot e_1|-\sqrt{\frac{2}{\pi}}\Big|\leq\delta_1\ll 1,\\&\Big|\frac{1}{m}\sum_{j=1}^m(a_j\cdot\hat{u})^2-1\Big|\leq\delta_1\ll 1,\quad\forall\hat{u}\in\mathbb{S}^{n-1}.\end{aligned}$$

Putting this into (2.2) gives

$$\begin{aligned}\partial_\rho f\leq&\frac{1}{m}\sum_{j=1}^m\Big(2\rho((a_j\cdot\hat{u})^2+2\beta)\Big)-\frac{1}{m}\sum_{j=1}^m 2\sqrt{\beta}|a_j\cdot e_1|\\\leq&2\rho(1+\delta_1+2\beta)-2\sqrt{\beta}\Big(\sqrt{\frac{2}{\pi}}-\delta_1\Big).\end{aligned}$$

Since $\sqrt{\frac{2}{\pi}}\approx 0.797885$, the desired result clearly follows by taking δ_1 sufficiently small. $\square$

The point $u=0$ needs to be treated with care since our loss function $f(u)$ is only Lipschitz at this point. To this end, we define the one-sided directional derivative of f along a direction $\xi\in\mathbb{S}^{n-1}$ as

$$D_\xi f(0)=\lim_{t\to 0^+}\frac{f(t\xi)}{t}. \tag{2.4}$$

It is easy to check that

$$D_\xi f(0)=-\frac{2}{m}\sum_{j=1}^m\sqrt{\beta+(a_j\cdot\xi)^2}|a_j\cdot e_1|.$$

Lemma 2.3 (The point $u=0$ is OK)**.** *For $m\gtrsim n$, with high probability it holds that*

$$D_\xi f(0)<-\sqrt{\beta},\quad\forall\xi\in\mathbb{S}^{n-1}.$$

Proof. Clearly with high probability and uniformly in $\xi \in \mathbb{S}^{n-1}$,

$$D_\xi f(0) \le -\frac{2}{m}\sum_{j=1}^m \sqrt{\beta}|a_j\cdot e_1| < -2\sqrt{\beta}\Big(\sqrt{\frac{2}{\pi}}-0.01\Big) < -\sqrt{\beta}.$$

So, we complete the proof. $\square$

In summary, we have the following theorem.

Theorem 2.2 (Non-vanishing gradient when $\|u\|_2 \le \frac{\sqrt{\beta}}{4(1+\beta)}$ or $\|u\|_2 \ge 3\sqrt{1+\beta}$)**.** *For $m \gtrsim n$, with high probability the following hold:*

1. *We have*

$$\begin{aligned} &\partial_\rho f < 0, \qquad \forall 0<\rho \le \frac{\sqrt{\beta}}{4(1+\beta)}, \qquad \forall \hat{u}\in\mathbb{S}^{n-1};\\ &\partial_\rho f > 0, \qquad \forall \rho \ge 3\sqrt{1+\beta}, \qquad \forall \hat{u}\in\mathbb{S}^{n-1}. \end{aligned}$$

2. *For $u=0$, we have*

$$D_\xi f(0) < -\sqrt{\beta}, \quad \forall \xi \in \mathbb{S}^{n-1},$$

where $D_\xi f(0)$ was defined in (2.4).

Proof. This follows from Lemmas 2.1, 2.2 and 2.3. $\square$

2.2 Analysis of the regime $\rho \sim 1$, $||\hat{u}\cdot e_1|-1| \ge \epsilon_0 > 0$

In this section we consider the regime $0<c_1<\rho<c_2<\infty$, $|\hat{u}\cdot e_1|<1-\epsilon_0$, where $0<\epsilon_0 \ll 1$. The choice of the constants c_1 and c_2 can be quite flexible. For example, we can take $c_1 = \frac{\sqrt{\beta}}{4(1+\beta)}$, $c_2 = 3(1+\beta)$. For this reason we write $\rho \sim 1$. To simplify the discussion, we need to employ a new coordinate system. Write

$$\begin{aligned} \hat{u} &= (\hat{u}\cdot e_1)e_1 + \tilde{u},\\ &= te_1 + \sqrt{1-t^2}e^\perp, \end{aligned}$$

where $e^\perp \in \mathbb{S}^{n-1}$ satisfies $e^\perp \cdot e_1 = 0$. Clearly in the regime $\rho \sim 1$, $|t|<1$, we have a smooth representation

$$u = \rho\hat{u} = \rho(te_1 + \sqrt{1-t^2}e^\perp) =: \psi(\rho, t, e^\perp).$$

The following pedestrian proposition shows that the landscape of a smooth function undergoes mild changes under smooth change of variables.

Proposition 2.1 (Criteria for no local minimum)**.** *In the regime $\rho \sim 1$, $|t|<1$, consider*

$$f(u)=f(\psi(\rho,t,e^{\perp}))=:g(\rho,t,e^{\perp}).$$

Then the following hold:

1. *If at some point $|\partial_t g|>0$, then $\|\nabla f\|_2>0$ at the corresponding point.*
2. *If at some point $\partial_{tt} g<0$, then either $\nabla f\neq 0$ at the corresponding point, or $\nabla f=0$ and f has a negative curvature at the corresponding point (i.e., a strict saddle).*

Proof. These easily follow from the formulae:

$$\begin{aligned}&\partial_t g=\nabla f\cdot\partial_t\psi,\\&\partial_{tt} g=\nabla f\cdot\partial_{tt}\psi+(\partial_t\psi)^T\nabla^2 f\partial_t\psi,\end{aligned}$$

where $\nabla^2 f=(\partial_{ij} f)$ denotes the Hessian matrix of f. $\square$

Proposition 2.1 allows us to simplify the computation greatly by looking only at the derivatives ∂_t and ∂_{tt}. We shall use these in the regime $|t|<1-\epsilon_0$, where $0<\epsilon_0\ll 1$. Now observe that

$$\frac{1}{2}\mathbb{E}f=\frac{1}{2}(1+2\beta)\rho^2+\frac{1}{2}-\rho\mathbb{E}\sqrt{\beta+(a\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a\cdot e_1)^2},$$

where $a\sim\mathcal{N}(0,\mathrm{I_n})$. Denote

$$X_1=a\cdot e_1 \quad\text{ and }\quad Y_1=a\cdot e^{\perp},$$

so that

$$a\cdot\hat{u}=tX_1+\sqrt{1-t^2}Y_1=:X_t.$$

We focus on the term

$$\begin{aligned}&\mathbb{E}\sqrt{\beta+(a\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a\cdot e_1)^2}\\=&\mathbb{E}\sqrt{\beta+X_t^2}\sqrt{\beta\rho^2+X_1^2}=:h_\infty(\rho,t).\end{aligned}$$

Lemma 2.4 (The limiting profile)**.** *For any $0<\eta_0\ll 1$, the following hold:*

1. $\sup_{|t|\le 1-\eta_0,\rho\sim 1}(|\partial_t h_\infty(\rho,t)|+|\partial_{tt} h_\infty(\rho,t)|+|\partial_{ttt} h_\infty(\rho,t)|)\lesssim 1$.
2. $|\partial_t h_\infty(\rho,t)|\gtrsim |t|$ *for* $0<|t|<1$, $\rho\sim 1$.

3. $\partial_{tt}h_\infty(\rho,t)\gtrsim 1$ *for* $|t|\ll 1$, $\rho\sim 1$.

Proof. See appendix. $\square$

Theorem 2.3 (The regime $\|u\|_2\sim 1$, $||\hat{u}\cdot e_1|-1|>\eta_0$ is fine)**.** *For any given* $0<\eta_0\ll 1$ *and* $0<c_1<c_2<\infty$, *if* $m\gtrsim n$, *then the following hold with high probability: In the regime* $c_1<\rho=\|u\|_2<c_2$, $||\hat{u}\cdot e_1|-1|>\eta_0$, *there are only two possibilities:*

1. $\|\nabla f(u)\|_2>0$;
2. $\nabla f(u)=0$, *and* f *has a negative directional curvature at this point.*

Proof. Denote

$$
\begin{aligned}
g(\rho,t,e^\perp)&=2\beta\rho^2+\rho^2\frac{1}{m}\sum_{j=1}^m(a_j\cdot\hat{u})^2-2\rho\frac{1}{m}\sum_{j=1}^m\sqrt{\beta+(a_j\cdot\hat{u})^2}\cdot\sqrt{\beta\rho^2+X_j^2}\\
&=:2\beta\rho^2+\rho^2h_0(\rho,t,e^\perp)-2\rho h(\rho,t,e^\perp),
\end{aligned}
$$

where $X_j=a_j\cdot e_1$, and

$$
a_j\cdot\hat{u}=tX_j+\sqrt{1-t^2}Y_j,\quad Y_j=a_j\cdot e^\perp.
$$

Clearly

$$
\begin{aligned}
&\partial_t g=\rho^2\partial_t h_0-2\rho\partial_t h;\\
&\partial_{tt} g=\rho^2\partial_{tt} h_0-2\rho\partial_{tt} h.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
h_0&=\frac{1}{m}\sum_{j=1}^m(tX_j+\sqrt{1-t^2}Y_j)^2\\
&=t^2\frac{1}{m}\sum_{j=1}^m X_j^2+2t\sqrt{1-t^2}\frac{1}{m}\sum_{j=1}^m X_jY_j+(1-t^2)\frac{1}{m}\sum_{j=1}^m Y_j^2.
\end{aligned}
$$

Clearly then for any small $\epsilon>0$ and $m\gtrsim n$, it holds with high probability that

$$
|\partial_t h_0-\mathbb{E}\partial_t h_0|+|\partial_{tt} h_0-\mathbb{E}\partial_{tt} h_0|\le\epsilon,\quad\forall|t|\le 1-\epsilon_0,\quad e^\perp\cdot e_1=0,\quad e^\perp\in\mathbb{S}^{n-1}.
$$

Note that we actually have $\mathbb{E}\partial_t h_0=0$ and $\mathbb{E}\partial_{tt}h_0=0$. By Lemmas A.4 and A.5, for any small $\epsilon>0$ and $m\gtrsim n$, it also holds with high probability that

$$
\begin{aligned}
&|\partial_t h-\mathbb{E}\partial_t h|\le\epsilon,\quad\forall|t|\le 1-\epsilon_0,\quad e^\perp\cdot e_1=0,\quad e^\perp\in\mathbb{S}^{n-1},\quad\forall c_1\le\rho\le c_2;\\
&\partial_{tt} h\ge\mathbb{E}\partial_{tt} h-\epsilon,\quad\forall|t|\le 1-\epsilon_0,\quad e^\perp\cdot e_1=0,\quad e^\perp\in\mathbb{S}^{n-1},\quad\forall c_1\le\rho\le c_2.
\end{aligned}
$$

We then obtain for small $\epsilon>0$, if $m\gtrsim n$, it holds with high probability that

$$|\partial_t g-\mathbb{E}\partial_t g|\le 2\epsilon;$$
$$\partial_{tt} g\le \mathbb{E}\partial_{tt} g+2\epsilon.$$

Clearly

$$\mathbb{E}\partial_t g=-2\rho\partial_t h_\infty(\rho,t);$$
$$\mathbb{E}\partial_{tt} g=-2\rho\partial_{tt} h_\infty(\rho,t).$$

By Lemma 2.4, we can take $t_0\ll 1$ such that

$$\begin{aligned} &\partial_{tt}h_\infty(\rho,t)\ge\epsilon_1>0, && \forall |t|\le t_0, && c_1\le\rho\le c_2;\\ &|\partial_t h_\infty(\rho,t)|\ge\epsilon_2>0, && \forall t_0\le|t|\le 1-\epsilon_0, && c_1\le\rho\le c_2. \end{aligned}$$

By taking $\epsilon>0$ sufficiently small, we can then guarantee that

$$\begin{aligned} &|\partial_t g|>\epsilon_3>0, && \forall |t|\le t_0, && c_1\le\rho\le c_2;\\ &\partial_{tt} g\le -\epsilon_4<0, && \forall t_0\le|t|\le 1-\epsilon_0, && c_1\le\rho\le c_2. \end{aligned}$$

The desired result then follows from Proposition 2.1. □

2.3 Localization of ρ, the regime $||\hat{u}\cdot e_1|-1|\ll 1$

In this section we shall localize ρ under the assumption that $||\hat{u}\cdot e_1|-1|\ll 1$, i.e., we shall show that if $||\hat{u}\cdot e_1|-1|\le\epsilon_0\ll 1$, then with high probability that $|\rho-1|\le\eta(\epsilon_0)\ll 1$. In the lemma below we assume $\rho\ge c_1$ since by Theorem 2.2 the regime $\rho\ll 1$ is already treated.

Lemma 2.5. *Let $0<\beta\le\frac{1}{4}$ and consider the regime $0<c_1\le\rho<1$. If $0<\eta_0\ll 1$ is sufficiently small, then for $m\gtrsim n$, it holds with high probability that*

$$\partial_\rho f<0,\quad \forall\rho\le 1-c(\eta_0),\quad \forall\hat{u}\in\mathbb{S}^{n-1}\quad \text{with}\quad ||\hat{u}\cdot e_1|-1|\le\eta_0,$$

where $c(\eta_0)\to 0$ as $\eta_0\to 0$.

Remark 2.2. In Theorem 2.4 we shall remove the constraint $0<\beta\le\frac{1}{4}$ and prove the result for all $0<\beta<\infty$.

Proof of Lemma 2.5. Recall

$$\begin{aligned}\frac{1}{2}\partial_\rho f=&\frac{1}{m}\sum_{j=1}^m\Big(\rho((a_j\cdot\hat{u})^2+2\beta)-\sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}\\ &-\frac{\beta\rho^2\sqrt{\beta+(a_j\cdot\hat{u})^2}}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}}\Big).\end{aligned}$$

Without loss of generality we assume

$$\|\hat{u}-e_1\|_2 \le r \ll 1.$$

The other case $\|\hat{u}+e_1\|_2 \ll 1$ is similar and therefore omitted. Note that

$$\left|\sqrt{\beta+(a_j\cdot\hat{u})^2}-\sqrt{\beta+(a_j\cdot e_1)^2}\right| = \left|\frac{(a_j\cdot(\hat{u}-e_1))(a_j\cdot(\hat{u}+e_1))}{\sqrt{\beta+(a_j\cdot\hat{u})^2}+\sqrt{\beta+(a_j\cdot e_1)^2}}\right| \le |a_j\cdot(\hat{u}-e_1)|.$$

It immediately gives

$$\begin{aligned}\frac{1}{2}\partial_\rho f \le & \frac{1}{m}\sum_{j=1}^m\Big(\rho((a_j\cdot\hat{u})^2+2\beta)-\sqrt{\beta+(a_j\cdot e_1)^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2} \\ & -\frac{\beta\rho^2\sqrt{\beta+(a_j\cdot e_1)^2}}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}}\Big)+H,\end{aligned}$$

where

$$H \lesssim \frac{1}{m}\sum_{j=1}^m |a_j\cdot(\hat{u}-e_1)|(1+|a_j\cdot e_1|) \lesssim \frac{1}{m}\sum_{j=1}^m\Big(a_j\cdot(\hat{u}-e_1))^2\cdot\frac{1}{2r}+r(1+(a_j\cdot e_1)^2)\Big).$$

Clearly it holds with high probability that

$$H \le B_1 r,$$

where $B_1>0$ is a constant. For $\rho\le 1$, we have

$$\begin{aligned}&-\sqrt{\beta+(a_j\cdot e_1)^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2} \le -(\beta\rho^2+(a_j\cdot e_1)^2); \\ &-\sqrt{\beta+(a_j\cdot e_1)^2}\cdot\frac{\beta\rho^2}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}} \le -\beta\rho^2.\end{aligned}$$

Then assuming $\rho<1$, it holds with high probability that

$$\frac{1}{2}\partial_\rho f \le \rho(1+2\beta)-(\beta\rho^2+1)-\beta\rho^2+B_1 r+\delta_1,$$

where $\delta_1>0$ is a small constant which accounts for the deviation from the mean value used in the Bernstein's inequality. For $0<\beta\le\frac{1}{4}$ (actually $0<\beta<\frac{1}{2}$ suffices) the desired conclusion then clearly follows by taking $\delta_1=\mathcal{O}(\eta_0)$ and $r=\mathcal{O}(\eta_0)$. $\square$

Lemma 2.6 ($\partial_{\rho\rho} f$ is good)**.** *We have almost surely it holds that*

$$\partial_{\rho\rho} f > 0, \quad \forall 0<\rho<\infty, \quad \forall \hat{u} \in \mathbb{S}^{n-1}.$$

Furthermore, for any fixed two constants $0<c_1<c_2<\infty$*, if* $m \gtrsim n$*, then it holds with high probability that*

$$\partial_{\rho\rho} f \geq \alpha > 0, \quad \forall c_1 \leq \rho \leq c_2, \quad \forall \hat{u} \in \mathbb{S}^{n-1},$$

where $\alpha>0$ *is a constant depending only on* (c_1,c_2,β)*.*

Proof. Recall

$$\frac{1}{2}\partial_\rho f = \frac{1}{m}\sum_{j=1}^m \Big(\rho((a_j\cdot\hat{u})^2+2\beta) - \sqrt{\beta+(a_j\cdot\hat{u})^2}\sqrt{\beta\rho^2+(a_j\cdot e_1)^2} - \frac{\beta\rho^2\sqrt{\beta+(a_j\cdot\hat{u})^2}}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}}\Big).$$

A simple calculation leads to

$$\begin{aligned}
&\frac{1}{2}\partial_{\rho\rho} f = \frac{1}{m}\sum_{j=1}^m \Big(((a_j\cdot\hat{u})^2+2\beta) - \frac{3\beta\rho\sqrt{\beta+(a_j\cdot\hat{u})^2}}{\sqrt{\beta\rho^2+(a_j\cdot e_1)^2}} + \frac{\beta^2\rho^3\sqrt{\beta+(a_j\cdot\hat{u})^2}}{(\beta\rho^2+(a_j\cdot e_1)^2)^{\frac{3}{2}}}\Big) \\
=&\frac{1}{m}\sum_{j=1}^m \left(((a_j\cdot\hat{u})^2+2\beta) - \sqrt{\beta+(a_j\cdot\hat{u})^2}\cdot\sqrt{\beta}\cdot\Big(3\left(\frac{a_j\cdot e_1}{\sqrt{\beta}\rho}\right)^2+2\Big)\cdot\Big(\left(\frac{a_j\cdot e_1}{\sqrt{\beta}\rho}\right)^2+1\Big)^{-\frac{3}{2}}\right).
\end{aligned}$$

For $0\leq x<\infty$, denote

$$h_0(x) = \frac{3x+2}{(1+x)^{\frac{3}{2}}}.$$

It is not difficult to check that

$$h_0(x) \leq 2, \quad \forall 0 \leq x < \infty, \tag{2.5}$$

and the equality holds if and only if $x=0$. Now define $h_1(x)=h_0(x^2)$. Then we can rewrite $\partial_{\rho\rho} f$ as

$$\frac{1}{2}\partial_{\rho\rho} f = \frac{1}{m}\sum_{j=1}^m \Big(\sqrt{\beta+(a_j\cdot\hat{u})^2}-\sqrt{\beta}\Big)^2 + \frac{1}{m}\sum_{j=1}^m \sqrt{\beta+(a_j\cdot\hat{u})^2}\cdot\sqrt{\beta}\cdot\Big(2-h_1\Big(\frac{a_j\cdot e_1}{\sqrt{\beta}\rho}\Big)\Big).$$

It then follows from (2.5) that

$$\frac{1}{2}\partial_{\rho\rho} f \geq \frac{\beta}{m}\sum_{j=1}^m \Big(2-h_1\Big(\frac{a_j\cdot e_1}{\sqrt{\beta}\rho}\Big)\Big) > 0$$

holds almost surely since the event $\bigcap_{j=1}^m \{a_j\cdot e_1=0\}$ has zero probability. By using the Bernstein's inequality, we have with high probability that

$$\frac{1}{m}\sum_{j=1}^m\Big(2-h_1(\frac{a_j\cdot e_1}{\sqrt{\beta}\rho})\Big)\gtrsim 1,\quad \forall c_1\le\rho\le c_2.$$

Thus

$$\partial_{\rho\rho}f\gtrsim 1,\quad \forall c_1\le\rho\le c_2,\quad \forall\hat{u}\in\mathbb{S}^{n-1}.$$

Thus, we complete the proof. $\square$

Theorem 2.4 (Localization of ρ when $||\hat{u}\cdot e_1|-1|\ll 1$)**.** *Consider the regime* $0<c_1\le\rho\le c_2$*. If* $0<\eta_0\ll 1$ *is sufficiently small, then for* $m\gtrsim n$*, it holds with high probability that*

$$\begin{aligned}
&\partial_\rho f<0, && \forall\rho\le 1-c(\eta_0), && \forall\hat{u}\in\mathbb{S}^{n-1} && \textit{with} && ||\hat{u}\cdot e_1|-1|\le\eta_0;\\
&\partial_\rho f>0, && \forall\rho\ge 1+c(\eta_0), && \forall\hat{u}\in\mathbb{S}^{n-1} && \textit{with} && ||\hat{u}\cdot e_1|-1|\le\eta_0;
\end{aligned}$$

where $c(\eta_0)\to 0$ *as* $\eta_0\to 0$.

Proof. We shall sketch the proof. We first consider the regime $\rho\ge 1$. Without loss of generality we assume $\|\hat{u}-e_1\|_2\le\eta_0$. The other case $\|\hat{u}+e_1\|_2\le\eta_0$ can be similarly treated.

First observe that

$$(\partial_\rho f)(\rho=1,\hat{u}=e_1)=0.$$

Then by a calculation similar to the estimate of H term in Lemma 2.5, we have with high probability that

$$|\partial_\rho h(\rho=1,\hat{u})|=c_1(\eta_0)\ll 1,\quad \forall|\hat{u}-e_1|\le\eta_0,$$

where $c_1(\eta_0)\to 0$ as $\eta_0\to 0$. Now by Lemma 2.6, it holds with high probability that

$$\partial_{\rho\rho}f\ge\alpha>0,\quad \forall c_1\le\rho\le c_2,\quad \forall\hat{u}\in\mathbb{S}^{n-1}.$$

It then implies that for $\rho\ge 1+\frac{2c_1(\eta_0)}{\alpha_0}$, we have

$$(\partial_\rho f)(\rho,\hat{u})\ge\alpha_0\cdot\frac{2c_1(\eta_0)}{\alpha_0}-c_1(\eta_0)=c_1(\eta_0)>0.$$

Redefining $c(\eta_0)$ suitably then yields the result. The argument for $\rho\le 1-c(\eta_0)$ is similar. We omit the details. $\square$

2.4 Strong convexity near the global minimizers $u=\pm e_1$: analysis of the limiting profile

In this section we shall show that in the small neighborhood of $u=\pm e_1$ where

$$||\hat{u}\cdot e_1|-1|\ll 1, \quad |\rho-1|\ll 1,$$

the Hessian of the expectation of the loss function must be strictly positive definite. In yet other words $\mathbb{E}f$ must be strictly convex in this neighborhood so that $u=\pm e_1$ are the unique minimizers. To this end consider

$$h(u)=\frac{1}{2}(\mathbb{E}f-1)=\frac{1}{2}(1+2\beta)|u|^2-\mathbb{E}\sqrt{\beta|u|^2+(a\cdot u)^2}\sqrt{\beta|u|^2+(a\cdot e_1)^2}, \tag{2.6}$$

where $a\sim\mathcal{N}(0,\mathrm{I}_n)$.

Theorem 2.5 (Strong convexity of $\mathbb{E}f$ when $\|u\pm e_1\|\ll 1$)**.** *Consider h defined by* (2.6)*. There exist $0<\epsilon_0\ll 1$ and a positive constant γ_1 such that the following hold:*

1. *If $\|u-e_1\|_2\le\epsilon_0$, then for any $\xi\in\mathbb{S}^{n-1}$ it holds*

$$\sum_{i,j=1}^n \xi_i\xi_j\mathbb{E}(\partial_i\partial_j h)(u)\ge\gamma_1>0.$$

2. *If $\|u+e_1\|_2\le\epsilon_0$, then for any $\xi\in\mathbb{S}^{n-1}$ it holds*

$$\sum_{i,j=1}^n \xi_i\xi_j\mathbb{E}(\partial_i\partial_j h)(u)\ge\gamma_1>0.$$

Proof. We shall only consider the case $\|u-e_1\|_2\ll 1$. The other case $\|u+e_1\|_2\ll 1$ is similar and therefore omitted. Note that

$$\|u-e_1\|_2^2=\|\rho\hat{u}-e_1\|_2^2=(\rho-1)^2+2\rho(1-t)\le\epsilon_0^2,$$

where $t=\hat{u}\cdot e_1$. Thus for $0<\epsilon_0\ll 1$, we have

$$|\rho-1|\le\epsilon_0, \quad 1-\epsilon_0^2\le t\le 1.$$

We now need to make a change of variable. The representation

$$u=te_1+\sqrt{1-t^2}e^{\perp}$$

is not so suitable since the derivatives blow up as $t\to 1-$. This is an artificial singularity due to the non-smoothness of the representation $\sqrt{1-t^2}$ as $t\to 1-$. To resolve this, we use a different representation (recall $1-\epsilon_0^2\le \hat{u}\cdot e_1\to 1$),

$$\hat{u}=\sqrt{1-s^2}e_1+se^{\perp},\quad e^{\perp}\cdot e_1=0,\quad e^{\perp}\in\mathbb{S}^{n-1},$$

where we assume $0\le s\ll 1$. Note that $s=\frac{|u'|}{\rho}$, and $u'=u-(u\cdot e_1)e_1=(0,u_2,\cdots,u_n)$.

To calculate $\partial^2 h$ we need to compute the Hessian expressed in the (ρ,s) coordinate. It is not difficult to check that by (2.6), the value of $h(u)$ depends only on (ρ,s). Thus by a slight abuse of notation we write $h=h(|u|,\frac{|u'|}{|u|})=h(\rho,s)$ (we denote $|u|=\|u\|_2$, $|u'|=\|u'\|_2$) and compute (below we assume $s>0$ so that $|u'|>0$)

$$\partial_i h=\partial_\rho h\frac{u_i}{\rho}+\partial_s h\cdot\Big(-\frac{u_i}{\rho^3}|u'|+1_{i\neq 1}\frac{1}{\rho}\cdot\frac{u_i}{|u'|}\Big)$$

and

$$\begin{aligned}\partial_{ij}h=&\partial_{\rho\rho}h\frac{u_iu_j}{\rho^2}+\frac{u_i}{\rho}\partial_{\rho s}h\cdot\Big(-\frac{u_j}{\rho^3}|u'|+\partial_\rho h\cdot\Big(\frac{\delta_{ij}}{\rho}-\frac{u_iu_j}{\rho^3}\Big)+\frac{u_j}{\rho}\partial_{\rho s}h\cdot\Big(-\frac{u_i}{\rho^3}|u'|+1_{i\neq 1}\frac{1}{\rho}\cdot\frac{u_i}{|u'|}\Big)\\&+1_{j\neq 1}\frac{1}{\rho}\cdot\frac{u_j}{|u'|}\Big)+\partial_{ss}h\cdot\Big(-\frac{u_i}{\rho^3}|u'|+1_{i\neq 1}\frac{1}{\rho}\cdot\frac{u_i}{|u'|}\Big)\cdot\Big(-\frac{u_j}{\rho^3}|u'|+1_{j\neq 1}\frac{1}{\rho}\cdot\frac{u_j}{|u'|}\Big)\\&+\partial_s h\cdot\Big(-\frac{\delta_{ij}}{\rho^3}|u'|+\frac{3u_iu_j}{\rho^5}|u'|-\frac{u_i}{\rho^3}\frac{u_j}{|u'|}1_{j\neq 1}-\frac{u_j}{\rho^3}\frac{u_i}{|u'|}1_{i\neq 1}+1_{i\neq 1}\frac{1}{\rho}\frac{\delta_{ij}}{|u'|}-1_{i\neq 1}1_{j\neq 1}\frac{1}{\rho}\cdot\frac{u_iu_j}{|u'|^3}\Big).\end{aligned}$$

Then denoting

$$a=\xi\cdot\hat{u},\quad b=\sum_{j\neq 1}\xi_j\cdot\frac{u_j}{|u'|},$$

we have

$$\begin{aligned}\sum_{i,j=1}^n\xi_i\xi_j\partial_{ij}h=&\partial_{\rho\rho}h\cdot a^2+2a\partial_{\rho s}h\cdot\Big(-\frac{as}{\rho}+\frac{b}{\rho}\Big)+\partial_\rho h\cdot\Big(\frac{|\xi|^2-|\xi\cdot\hat{u}|^2}{\rho}\Big)+\partial_{ss}h\cdot\Big(-\frac{as}{\rho}+\frac{b}{\rho}\Big)^2\\&+\partial_s h\cdot\Big(-\frac{|\xi|^2}{\rho^2}s+3\frac{a^2s}{\rho^2}-2\frac{ab}{\rho^2}+\frac{|\xi'|^2}{\rho|u'|}-\frac{b^2}{\rho|u'|}\Big)\\=&\partial_{\rho\rho}h\cdot a^2+2a\partial_{\rho s}h\cdot\Big(-\frac{as}{\rho}+\frac{b}{\rho}\Big)\quad+\partial_\rho h\cdot\Big(\frac{|\xi|^2-|\xi\cdot\hat{u}|^2}{\rho}\Big)\\&+\partial_{ss}h\cdot\Big(\frac{a^2s^2-2abs}{\rho^2}\Big)+\Big(\partial_{ss}h-\frac{1}{s}\partial_s h\Big)\frac{b^2}{\rho^2}+\partial_s h\cdot\Big(-\frac{|\xi|^2}{\rho^2}s+3\frac{a^2s}{\rho^2}-2\frac{ab}{\rho^2}+\frac{|\xi'|^2}{\rho^2 s}\Big).\end{aligned}$$

We should point it out that, in the above computation, one does not need to worry about the formal singularity caused by $\frac{1}{s}$. Since $\partial_s h(\rho,s=0)=0$, we write

$$(\partial_s h)(\rho,s)\cdot\frac{1}{s}=\frac{(\partial_s h)(\rho,s)-(\partial_s h)(\rho,0)}{s}=\int_0^1(\partial_{ss}h)(\rho,\theta s)d\theta,\quad s>0.$$

In particular we have

$$\begin{aligned}
&\lim_{s\to 0^+}(\partial_s h)(\rho,s)\cdot\frac{1}{s}=(\partial_{ss}h)(\rho,0);\\
&\Big|(\partial_{ss}h)(\rho,s)-\frac{1}{s}\partial_s h_1(\rho,s)\Big|=\mathcal{O}(s)\to 0 \quad \text{as} \quad s\to 0.
\end{aligned}$$

By using this observation and Lemma A.7, we obtain

$$\begin{aligned}
\sum_{i,j=1}^n \xi_i\xi_j(\partial_{ij}h)(e_1)=&(\partial_{\rho\rho}h)(1,0)\cdot a^2\Big|_{a=\xi_1}+(\partial_{ss}h)(1,0)\cdot|\xi'|^2\\
\ge&\gamma_0\cdot|\xi|^2, \quad \forall \xi\in\mathbb{S}^{n-1},
\end{aligned}$$

where $\gamma_0>0$ is a constant. Now for $\|u-e_1\|_2\ll 1$, by using Lemma A.6 and Lemma A.7, we have

$$\begin{aligned}
&\Big|\sum_{i,j=1}^n \xi_i\xi_j\Big((\partial_{ij}h)(e_1)-(\partial_{ij}h)(u)\Big)\Big|\\
\lesssim&|(\partial_{\rho\rho}h)(\rho,s)-(\partial_{\rho\rho}h)(1,0)|+|(\partial_{\rho\rho}h)(\rho,s)|\cdot|(\xi\cdot\hat{u})^2-(\xi\cdot e_1)^2|\\
&+|(\partial_{\rho s}h)(\rho,s)|\cdot(1+s)+|(\partial_\rho h)(\rho,s)|+|\partial_{ss}h(\rho,s)|\cdot(s+s^2)\\
&+|\partial_s h(\rho,s)|+\Big|\frac{1}{\rho^2}\int_0^1(\partial_{ss}h)(\rho,\theta s)d\theta-(\partial_{ss}h)(1,0)\Big|\\
&+\Big|(\partial_{ss}h)(\rho,s)-(\partial_s h)(\rho,s)\cdot\frac{1}{s}\Big|\cdot\frac{b^2}{\rho^2}\\
\lesssim&\mathcal{O}(|\rho-1|+|s|+\|\hat{u}-e_1\|_2).
\end{aligned}$$

It follows that if $\|u-e_1\|_2$ is sufficiently small, we then have

$$\sum_{i,j=1}^n \xi_i\xi_j(\partial_{ij}h)(u)\ge\frac{\gamma_0}{2}|\xi|^2, \quad \forall\xi\in\mathbb{S}^{n-1}.$$

This completes the proof. □

2.5 Near the global minimizer: strong convexity

In this section we show strong convexity of the loss function $f(u)$ near the global minimizer $u=\pm e_1$.

Theorem 2.6 (Strong convexity near the global minimizer)**.** *There exist $0<\epsilon_0\ll 1$ and positive constants β_1,γ such that if $m\gtrsim n$, then the following hold with probability at least $1-\frac{\beta_1}{m^2}$:*

1. If $\|u-e_1\|_2 \le \epsilon_0$*, then*

$$\sum_{i,j=1}^n \xi_i \xi_j (\partial_{ij} f)(u) \ge \gamma > 0, \quad \forall \xi \in \mathbb{S}^{n-1}.$$

2. If $\|u+e_1\|_2 \le \epsilon_0$*, then*

$$\sum_{i,j=1}^n \xi_i \xi_j (\partial_{ij} f)(u) \ge \gamma > 0, \quad \forall \xi \in \mathbb{S}^{n-1}.$$

In other words, $f(u)$ *is strongly convex in a sufficiently small neighborhood of* $\pm e_1$*.*

Proof. Recall

$$f(u) = 2f_0(u) + \frac{1}{m}\sum_{k=1}^m ((a_k\cdot u)^2 + 2\beta|u|^2),$$

where

$$f_0(u) = -\frac{1}{m}\sum_{k=1}^m \sqrt{\beta|u|^2+(a_k\cdot u)^2}\cdot\sqrt{\beta|u|^2+(a_k\cdot e_1)^2}.$$

Clearly

$$\sum_{i,j=1}^n \xi_i\xi_j \partial_{ij} f(u) = 2\Big(\frac{1}{m}\sum_{k=1}^m |a_k\cdot\xi|^2\Big) + 4\beta|\xi|^2 + 2\sum_{i,j=1}^n \xi_i\xi_j\partial_{ij} f_0(u).$$

Obviously we have for $m \gtrsim n$, it holds with high probability that

$$\Big|\frac{1}{m}\sum_{k=1}^m |a_k\cdot\xi|^2 - 1\Big| \le \frac{\epsilon}{100}, \quad \forall \xi \in \mathbb{S}^{n-1}.$$

By Lemma A.10, we have

$$\Big|\sum_{i,j=1}^n \xi_i\xi_j(\partial_{ij} f_0(u) - \mathbb{E}\partial_{ij} f_0(u))\Big| \le \frac{\epsilon}{100}, \quad \forall \xi\in\mathbb{S}^{n-1}, \quad \forall \frac{1}{3} \le \|u\|_2 \le 3.$$

Thus we have

$$\Big|\sum_{i,j=1}^n \xi_i\xi_j(\partial_{ij} f(u) - \mathbb{E}\partial_{ij} f(u))\Big| \le \frac{\epsilon}{100}, \quad \forall \xi\in\mathbb{S}^{n-1}, \quad \forall \frac{1}{3} \le \|u\|_2 \le 3.$$

The desired result then follows from Theorem 2.5 by taking $\epsilon > 0$ sufficiently small. This completes the proof. $\square$

We now complete the proof of the main theorem.

Proof of Theorem 2.1. We proceed in several steps.

1. By Theorem 2.2, we see that with high probability the function $f(u)$ has non-vanishing gradient in the regimes

$$0<\|u\|_2\le\frac{\sqrt{\beta}}{4(1+\beta)}=c_1$$

and

$$\|u\|_2\ge 3(1+\beta)=c_2.$$

Moreover at the point $u=0$, we have the directional gradient is strictly less than $-\sqrt{\beta}$ along any direction $\xi\in\mathbb{S}^{n-1}$.

2. By Theorem 2.6, there exists $\epsilon_0>0$ sufficiently small, such that with probability at least $1-\mathcal{O}(m^{-2})$, $f(u)$ is strongly convex in the neighborhood $\|u\pm e_1\|_2\le\epsilon_0$.

3. By Theorem 2.4, we have that with high probability

$$\|\nabla f\|_2>0,$$

if $|\rho-1|\ge c(\eta_0)$ and $||\hat{u}\cdot e_1|-1|\le\eta_0$. Here we recall $\rho=\|u\|_2$ and $u=\rho\hat{u}$. Observe that

$$\|u\pm e_1\|_2^2=(\rho-1)^2+2\rho(1\pm\hat{u}\cdot e_1).$$

By taking $\eta_0=\epsilon_0^2/100$, we see that $\|u\pm e_1\|_2>\epsilon_0$, $||\hat{u}\cdot e_1|-1|\le\eta_0$ must imply

$$|\rho-1|>\frac{\epsilon_0}{10}.$$

Thus it remains for us to treat the regime $|||\hat{u}\cdot e_1|-1|>\eta_0$, $c_1\le\|u\|_2\le c_2$.

4. In the regime $||\hat{u}\cdot e_1|-1|>\eta_0$, $\|u\|_2\sim 1$, we have by Theorem 2.3, with high probability it holds that either the function has a non-vanishing gradient at the point u, or the gradient vanishes at u, but f has a negative directional curvature at this point.

This completes the proof. $\square$

3 Perturbed amplitude model II

In this section, we introduce the second perturbed amplitude model for solving phase retrieval problem and consider the global landscape of it. Specifically, we consider the following empirical loss for some parameter $\beta>0$,

$$f(u)=\frac{1}{m}\sum_{j=1}^{m}\Big(\sqrt{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot x)^2}-\sqrt{\beta|u|^2+2(a_j\cdot x)^2}\,\Big)^2. \tag{3.1}$$

Theorem 3.1. *Let $0<\beta<\infty$. Assume $\{a_i\}_{i=1}^m$ are i.i.d. standard Gaussian random vectors and $x\neq 0$. There exist positive constants c, C depending only on β, such that if $m\geq Cn$, then with probability at least $1-e^{-cm}$ the loss function $f=f(u)$ defined by (3.1) has no spurious local minimizers. The only local minimizer is $\pm x$, and the loss function is strongly convex in a neighborhood of $\pm x$. The point $u=0$ is a local maximum point with strictly negative-definite Hessian. All other critical points are strict saddles, i.e., each saddle point has a neighborhood where the function has negative directional curvature.*

Remark 3.1. One should note that the set $\bigcup_{j=1}^m\{a_j\cdot x=0\}$ has measure zero. Therefore for a typical realization, $a_j\cdot x$ is always non-zero for all j and the function

$$\tilde{f}_j(y)=\sqrt{y^2+(a_j\cdot x)^2}$$

is smooth. In particular, we can compute (for each realization) the derivatives of the summands in (3.1) without any problem.

Remark 3.2. Thanks to the regularization term $(a_j\cdot x)^2$, our new model (3.1) enjoys a better probability concentration bound $1-e^{-cm}$ than the model (2.1) where the weaker probability concentration $1-\mathcal{O}(m^{-2})$ is proved.

Without loss of generality we shall assume $x=e_1$ throughout the rest of the proof.

3.1 The regimes $\|u\|_2\ll 1$ and $\|u\|_2\gg 1$ are fine

Write $u=\rho\hat{u}$ where $\hat{u}\in\mathbb{S}^{n-1}$. Then

$$\begin{aligned}&\Big(\sqrt{\beta|u|^2+(a_j\cdot u)^2+|a_j\cdot e_1|^2}-\sqrt{\beta|u|^2+2(a_j\cdot e_1)^2}\,\Big)^2\\ =&\rho^2((a_j\cdot\hat{u})^2+2\beta)+3(a_j\cdot e_1)^2-2\sqrt{\beta\rho^2+\rho^2(a_j\cdot\hat{u})^2+(a_j\cdot e_1)^2}\sqrt{\beta\rho^2+2(a_j\cdot e_1)^2}.\end{aligned}$$

Thus, the derivative of f is

$$\begin{aligned}\partial_\rho f =&2\rho\frac{1}{m}\sum_{j=1}^m\Big(((a_j\cdot\hat{u})^2+2\beta)-\frac{\beta+(a_j\cdot\hat{u})^2}{\sqrt{\beta\rho^2+\rho^2(a_j\cdot\hat{u})^2+(a_j\cdot e_1)^2}}\sqrt{\beta\rho^2+2(a_j\cdot e_1)^2}\\&-\sqrt{\beta\rho^2+\rho^2(a_j\cdot\hat{u})^2+(a_j\cdot e_1)^2}\frac{\beta}{\sqrt{\beta\rho^2+2(a_j\cdot e_1)^2}}\Big).\end{aligned}\tag{3.2}$$

Lemma 3.1 (The regime $\rho\gg 1$ is OK)**.** *There exist constants $R_1=R_1(\beta)>0$, $d_1=d_1(\beta)>0$ such that the following hold: For $m\gtrsim n$, with high probability it holds that*

$$\partial_\rho f\geq d_1\rho,\quad \forall \rho\geq R_1,\quad \forall\hat{u}\in\mathbb{S}^{n-1}.$$

Proof. We only sketch the proof. Denote $X_j=a_j\cdot e_1$ and $Z_j=a_j\cdot\hat{u}$. We next gives several estimation bounds for the terms of $\partial_\rho f$. We first establish an upper bound for the second term. Before proceeding, observe that

$$\begin{aligned}&\frac{\beta\rho^2+2X_j^2}{(\beta+Z_j^2)\rho^2+X_j^2}-\frac{\beta}{\beta+Z_j^2}=\frac{\beta X_j^2+2X_j^2Z_j^2}{\Big((\beta+Z_j^2)\rho^2+X_j^2\Big)\cdot(\beta+Z_j^2)}\\&\leq\frac{2X_j^2}{(\beta+Z_j^2)\rho^2+X_j^2}\leq\frac{1}{\rho^2}\cdot\frac{2X_j^2}{\beta+Z_j^2},\end{aligned}$$

which means

$$\begin{aligned}(\beta+Z_j^2)\cdot\frac{\sqrt{\beta\rho^2+2X_j^2}}{\sqrt{\rho^2(\beta+Z_j^2)+X_j^2}}&\leq(\beta+Z_j^2)\Big(\sqrt{\frac{\beta}{\beta+Z_j^2}}+\frac{1}{\rho}\frac{\sqrt{2}|X_j|}{\sqrt{\beta+Z_j^2}}\Big)\\&\leq\sqrt{\beta}\sqrt{\beta+Z_j^2}+\frac{1}{\rho}\sqrt{2}|X_j|\sqrt{\beta+Z_j^2},\end{aligned}$$

where we use the fact $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for any positive number a,b in the first inequality. For the third term, it is easy to see that

$$\sqrt{\rho^2(\beta+Z_j^2)+X_j^2}\cdot\frac{\beta}{\sqrt{\beta\rho^2+2X_j^2}}\leq\sqrt{\beta}\cdot\sqrt{\beta+Z_j^2}.$$

Putting the above two estimators into (3.2), we get

$$\begin{aligned}\frac{1}{2\rho}\partial_\rho f&\geq\frac{1}{m}\sum_{j=1}^m\Big(Z_j^2+2\beta-2\sqrt{\beta}\sqrt{\beta+Z_j^2}\Big)-\frac{1}{\rho}\sqrt{2}\frac{1}{m}\sum_{j=1}^m|X_j|\cdot\sqrt{\beta+Z_j^2}\\&\geq\frac{1}{m}\sum_{j=1}^m\Big(\sqrt{Z_j^2+\beta}-\sqrt{\beta}\Big)^2-\frac{1}{\rho}\cdot\frac{1}{m}\sum_{j=1}^m(X_j^2+Z_j^2+\beta).\end{aligned}$$

By Bernstein's inequality and simple union bound arguments, we clearly have with high probability,

$$\begin{aligned}
&\Big|\frac{1}{m}\sum_{j=1}^{m}\Big(\sqrt{Z_j^2+\beta}-\sqrt{\beta}\Big)^2-\text{mean}\Big|\le\epsilon, && \forall\hat{u}\in\mathbb{S}^{n-1};\\
&\frac{1}{m}\sum_{j=1}^{m}(X_j^2+Z_j^2+\beta)\le 3+\beta, && \forall\hat{u}\in\mathbb{S}^{n-1}.
\end{aligned}$$

The desired result then clearly follows. $\square$

Lemma 3.2 (The regime $\|u\|_2\ll 1$ is OK)**.** *There exist constants $R_2=R_2(\beta)>0$, $d_2=d_2(\beta)>0$ such that the following hold: For $m\gtrsim n$, with high probability it holds that*

$$\partial_\rho f\le -d_2\rho<0,\quad \forall 0<\rho\le R_2,\quad \forall\hat{u}\in\mathbb{S}^{n-1}.$$

Moreover, at $u=0$, we have $\nabla f(0)=0$, and

$$\sum_{k,l=1}^{n}\xi_k\xi_l(\partial_{kl}f)(0)\le -d_2\|\xi\|_2^2,\quad \forall\xi\in\mathbb{S}^{n-1}.$$

In yet other words, $u=0$ is a strict local maximum point with strictly negative definite Hessian.

Proof. We only sketch the proof. Again denote $X_j=a_j\cdot e_1$ and $Z_j=a_j\cdot\hat{u}$. Observe that

$$\frac{\beta\rho^2+2X_j^2}{\rho^2(\beta+Z_j^2)+X_j^2}=2-\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}$$

and

$$(\beta+Z_j^2)\sqrt{\frac{\beta\rho^2+2X_j^2}{\rho^2(\beta+Z_j^2)+X_j^2}}\ge\sqrt{2}(\beta+Z_j^2)-(\beta+Z_j^2)\sqrt{\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}}.$$

On the other hand,

$$\frac{\rho^2(\beta+Z_j^2)+X_j^2}{\beta\rho^2+2X_j^2}-\frac{1}{2}=\frac{1}{2}\cdot\frac{\rho^2(\beta+2Z_j^2)}{\beta\rho^2+2X_j^2}\ge 0.$$

Thus

$$\begin{aligned}\frac{1}{2\rho}\partial_\rho f \le & \frac{1}{m}\sum_{j=1}^m \Big(Z_j^2+2\beta-\sqrt{2}\cdot(\beta+Z_j^2)-\frac{1}{\sqrt{2}}\beta\Big)\\ &+\frac{1}{m}\sum_{j=1}^m(\beta+Z_j^2)\sqrt{\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}}.\end{aligned}$$

Since

$$Z_j^2+2\beta-\sqrt{2}\cdot(\beta+Z_j^2)-\frac{1}{\sqrt{2}}\beta=-(\sqrt{2}-1)Z_j^2-\Big(\sqrt{2}+\frac{1}{\sqrt{2}}-2\Big)\beta,$$

the first summand clearly gives a nontrivial negative lower bound. The desired result then follows from Lemma B.1. We note that the result for $u=0$ follows by taking $u=t\xi$ and re-run the above argument taking $t\to 0+$. $\square$

Theorem 3.2 (The regimes $\|u\|_2\ll 1$ and $\|u\|_2\gg 1$ are OK)**.** *For $m\gtrsim n$, with high probability the following hold:*

1. *We have*

$$\begin{aligned}&\partial_\rho f\ge d_1\rho>0, &&\forall \rho\ge R_1, &&\forall \hat{u}\in\mathbb{S}^{n-1};\\ &\partial_\rho f\le -d_2\rho<0, &&\forall 0<\rho\le R_2, &&\forall \hat{u}\in\mathbb{S}^{n-1},\end{aligned}$$

   where d_1, d_2, R_1, R_2 are constants depending only on β.

2. *The point $u=0$ is a local maximum point with strictly negative-definite Hessian,*

$$\sum_{k,l=1}^n \xi_k\xi_l(\partial_{kl}f)(0)\le -d_2<0,\quad \forall \xi\in\mathbb{S}^{n-1}.$$

Proof. This follows from Lemmas 3.1 and 3.2. $\square$

Theorem 3.3 (The regime $\|u\|_2\sim 1$, $||\hat{u}\cdot e_1|-1|\le\epsilon_0$, $|\|u\|_2-1|\ge c(\epsilon_0)$ is OK)**.** *Let R_1, R_2 be the same as in Lemmas 3.1 and 3.2. Let $0<\epsilon_0\ll 1$ be given and consider the regime*

$$\Big||\hat{u}\cdot e_1|-1\Big|\le\epsilon_0 \quad \text{with} \quad R_1\le\|u\|_2\le R_2.$$

There exists a constant $c(\epsilon_0)>0$ ($c(\epsilon_0)$ also depends on β but we suppress this dependence) which tends to zero as $\epsilon_0\to 0$ such that the following hold: For $m\gtrsim n$, with high probability it holds that

$$\begin{aligned}&\partial_\rho f<0, &&\forall R_2\le\rho\le 1-c(\epsilon_0), &&\forall\hat{u}\in\mathbb{S}^{n-1} &&\text{with } ||\hat{u}\cdot e_1|-1|\le\epsilon_0;\\ &\partial_\rho f>0, &&\forall 1+c(\epsilon_0)\le\rho\le R_1, &&\forall\hat{u}\in\mathbb{S}^{n-1} &&\text{with } ||\hat{u}\cdot e_1|-1|\le\epsilon_0.\end{aligned}$$

Proof. We shall work with the variable $R=\rho^2$. Write

$$f(u)=\frac{1}{m}\sum_{j=1}^{m}F(\rho^2,(a_j\cdot\hat{u})^2,(a_j\cdot e_1)^2),$$

where

$$F(R,s,t)=R(s+2\beta)+3t-2\sqrt{R(\beta+s)+t}\sqrt{\beta R+2t}.$$

Observe that

$$\begin{aligned}\partial_R F&=s+2\beta-\Big(\beta\sqrt{\frac{R(\beta+s)+t}{\beta R+2t}}+(\beta+s)\sqrt{\frac{\beta R+2t}{R(\beta+s)+t}}\Big)\\&=s+2\beta-F_0(\sqrt{z(R,s,t)}),\end{aligned}\tag{3.3}$$

where

$$F_0(y)=\beta y+(\beta+s)y^{-1},\quad z(R,s,t)=\frac{R(\beta+s)+t}{\beta R+2t}.$$

Note that $F_0'(y)<0$ for any $0<y<\sqrt{\frac{\beta+s}{\beta}}$. It is easy to check that for $t>0$, $s\geq 0$ (note that $t=(a_j\cdot e_1)^2>0$ for all j almost surely)

$$z(R,s,t)<\frac{\beta+s}{\beta}.$$

Furthermore, since

$$z(R,s,t)=\frac{\beta+s}{\beta}-\frac{t\cdot\frac{\beta+2s}{\beta}}{\beta R+2t},$$

we have $\partial_R z(R,s,t)>0$ for $R>0$, $t>0$, $s\geq 0$. Thus

$$\partial_{RR}F>0,\quad \forall R>0,\quad \forall s\geq 0,\quad t>0.\tag{3.4}$$

On the other hand, by directly using (3.3), it is not difficult to check that for $R\sim 1$,

$$|(\partial_R F)(R,s,t)-(\partial_R F)(R,t,t)|\lesssim|s-t|.\tag{3.5}$$

Also observe that for $R\sim 1$, we have

$$\begin{aligned}&z(R,t,t)\sim 1,\quad \partial_R z(R,t,t)\sim\frac{t}{1+t};\\&-F_0'(\sqrt{z(R,t,t)})=-\beta+\frac{\beta+t}{z(R,t,t)}=\frac{1}{z(R,t,t)}\cdot\frac{t(\beta+2t)}{\beta R+2t}\sim t;\\&\partial_{RR}F(R,t,t)=-F_0'(\sqrt{z(R,t,t)})\frac{1}{2}z(R,t,t)^{-\frac{1}{2}}\cdot\partial_R z(R,t,t)\sim\frac{t^2}{1+t}.\end{aligned}\tag{3.6}$$

Note that $z(1,t,t)=1$ and $\partial_R F(1,t,t)=0$. By using (3.4), (3.5) and (3.6), we obtain for $R\geq 1+\eta_0$ ($0<\eta_0\ll 1$ will be specified later)

$$\begin{aligned}(\partial_R F)(R,s,t)>&(\partial_R F)(1+\eta_0,s,t)\\ \geq&(\partial_R F)(1+\eta_0,t,t)-B_1|s-t|\\ \geq&B_2\cdot\frac{t^2}{1+t}\cdot\eta_0-B_1|s-t|,\end{aligned}$$

where $B_1>0$, $B_2>0$ are constants depending only on β. Consequently we have for $R\geq 1+\eta_0$,

$$\partial_R f\geq B_2\eta_0\frac{1}{m}\sum_{j=1}^m\frac{(a_j\cdot e_1)^4}{1+(a_j\cdot e_1)^2}-B_1\frac{1}{m}\sum_{j=1}^m|(a_j\cdot\hat{u})^2-(a_j\cdot e_1)^2|.$$

Clearly with high probability,

$$\begin{aligned}&\frac{1}{m}\sum_{j=1}^m\frac{(a_j\cdot e_1)^4}{1+(a_j\cdot e_1)^2}\geq B_3>0,\\ &\frac{1}{m}\sum_{j=1}^m|(a_j\cdot\hat{u})^2-(a_j\cdot e_1)^2|\leq\frac{1}{m}\sum_{j=1}^m|a_j\cdot(\hat{u}-e_1)|\cdot|a_j\cdot(\hat{u}+e_1)|\\ &\leq B_4\min\{\|\hat{u}-e_1\|_2,\|\hat{u}+e_1\|_2\},\quad\forall\hat{u}\in\mathbb{S}^{n-1},\end{aligned}$$

where $B_3>0$, $B_4>0$ are absolute constants. Clearly then for $R\geq 1+\eta_0$, we have (below $B_5>0$, $B_6>0$ are constants depending only on β)

$$\partial_R f\geq B_5\eta_0-B_6\sqrt{1-|\hat{u}\cdot e_1|}>0,$$

if η_0 is chosen suitably small. The case for $R\leq 1-\eta_0$ is similar. We omit the details. □

3.2 Analysis of the regime $\rho\sim 1$, $||\hat{u}\cdot e_1|-1|\geq\epsilon_0>0$

In this section we consider the regime $\rho\sim 1$, $|\hat{u}\cdot e_1|<1-\epsilon_0$, where $0<\epsilon_0\ll 1$. To simplify the discussion, we use the coordinate system

$$\begin{aligned}\hat{u}=&(\hat{u}\cdot e_1)e_1+\tilde{u}\\ =&te_1+\sqrt{1-t^2}e^{\perp},\end{aligned}$$

where $e^{\perp}\in\mathbb{S}^{n-1}$ satisfies $e^{\perp}\cdot e_1=0$. Clearly in the regime $\rho\sim 1$, $|t|<1$, we have a smooth representation

$$u=\rho\hat{u}=\rho\cdot(te_1+\sqrt{1-t^2}e^{\perp})=\psi(\rho,t,e^{\perp}).$$

By Proposition 2.1, we can simplify the computation by examining only at the derivatives ∂_t and ∂_{tt}. We shall use these in the regime $|t|<1-\epsilon_0$ where $0<\epsilon_0\ll 1$. Now observe

$$\frac{1}{2}\mathbb{E}f=\frac{1}{2}(1+2\beta)\rho^2+\frac{3}{2}-\mathbb{E}\sqrt{\beta\rho^2+\rho^2(a\cdot\hat{u})^2+(a\cdot e_1)^2}\sqrt{\beta\rho^2+2(a\cdot e_1)^2},$$

where $a\sim\mathcal{N}(0,\mathrm{I_n})$. Denote

$$X_1=a\cdot e_1 \quad \text{and} \quad Y_1=a\cdot e^{\perp},$$

so that $a\cdot\hat{u}=tX_1+\sqrt{1-t^2}Y_1=:X_t$. We focus on the term

$$\begin{aligned}&\mathbb{E}\sqrt{\beta\rho^2+\rho^2(a\cdot\hat{u})^2+(a\cdot e_1)^2}\sqrt{\beta\rho^2+2(a\cdot e_1)^2}\\=&\mathbb{E}\sqrt{\beta\rho^2+\rho^2X_t^2+X_1^2}\sqrt{\beta\rho^2+2X_1^2}=:h_\infty(\rho,t).\end{aligned}$$

Lemma 3.3 (The limiting profile)**.** *For any $0<\eta_0\ll 1$, the following hold:*

1. $\sup_{|t|\le 1-\eta_0,\rho\sim 1}(|\partial_t h_\infty(\rho,t)|+|\partial_{tt}h_\infty(\rho,t)|+|\partial_{ttt}h_\infty(\rho,t)|)\lesssim 1$.
2. $|\partial_t h_\infty(\rho,t)|\gtrsim|t|$ *for* $0<|t|<1$, $\rho\sim 1$.
3. $\partial_{tt}h_\infty(\rho,t)\gtrsim 1$ *for* $|t|\ll 1$, $\rho\sim 1$.

Proof. See appendix. $\square$

Theorem 3.4 (The regime $\|u\|_2\sim 1$, $||\hat{u}\cdot e_1|-1|>\eta_0$ is fine)**.** *For any given $0<\eta_0\ll 1$ and $0<c_1<c_2<\infty$, if $m\gtrsim n$, then the following hold with high probability: In the regime $c_1<\rho=\|u\|_2<c_2$, $||\hat{u}\cdot e_1|-1|>\epsilon_0$, there are only two possibilities:*

1. $\|\nabla f(u)\|_2>0$;
2. $\nabla f(u)=0$, *and f has a negative directional curvature at this point.*

Proof. Denote $X_j=a_j\cdot e_1$ and

$$\begin{aligned}g(\rho,t,e^{\perp})=&2\beta\rho^2+\rho^2\frac{1}{m}\sum_{j=1}^m(a_j\cdot\hat{u})^2-2\frac{1}{m}\sum_{j=1}^m\sqrt{\beta\rho^2+\rho^2(a_j\cdot\hat{u})^2+X_j^2}\sqrt{\beta\rho^2+2X_j^2}\\=:&2\beta\rho^2+\rho^2h_0(\rho,t,e^{\perp})-2h(\rho,t,e^{\perp}),\end{aligned}$$

where

$$a_j\cdot\hat{u}=tX_j+\sqrt{1-t^2}Y_j,\quad X_j=a_j\cdot e_1,\quad Y_j=a_j\cdot e^{\perp}.$$

By the same argument as in the proof of Theorem 2.3, we have for any small $\epsilon>0$ and $m\gtrsim n$, it holds with high probability that

$$|\partial_t h_0-\mathbb{E}\partial_t h_0|+|\partial_{tt}h_0-\mathbb{E}\partial_{tt}h_0|\le\epsilon,\quad\forall|t|\le 1-\epsilon_0,\quad e^{\perp}\cdot e_1=0,\quad e^{\perp}\in\mathbb{S}^{n-1}.$$

Note also $\mathbb{E}\partial_t h_0=0$ and $\mathbb{E}\partial_{tt}h_0=0$. By Lemma B.4, for any small $\epsilon>0$ and $m\gtrsim n$, it also holds with high probability that

$$|\partial_t h-\mathbb{E}\partial_t h|+|\partial_{tt}h-\mathbb{E}\partial_{tt}h|\le\epsilon,\quad\forall|t|\le 1-\epsilon_0,\quad e^{\perp}\cdot e_1=0,\quad e^{\perp}\in\mathbb{S}^{n-1},\quad\forall c_1\le\rho\le c_2.$$

We then obtain for small $\epsilon>0$, if $m\gtrsim n$, it holds with high probability that

$$\begin{aligned}&|\partial_t g-\mathbb{E}\partial_t g|\le 2\epsilon;\\&\partial_{tt}g\le\mathbb{E}\partial_{tt}g+2\epsilon.\end{aligned}$$

Clearly

$$\begin{aligned}&\mathbb{E}\partial_t g=-\partial_t h_\infty(\rho,t);\\&\mathbb{E}\partial_{tt}g=-\partial_{tt}h_\infty(\rho,t).\end{aligned}$$

By Lemma 3.3, we can take $t_0\ll 1$ such that

$$\begin{aligned}&\partial_{tt}h_\infty(\rho,t)\ge\epsilon_1>0, &&\forall|t|\le t_0, &&c_1\le\rho\le c_2;\\&|\partial_t h_\infty(\rho,t)|\ge\epsilon_2>0, &&\forall t_0\le|t|\le 1-\epsilon_0, &&c_1\le\rho\le c_2.\end{aligned}$$

By taking $\epsilon>0$ sufficiently small, we can then guarantee that

$$\begin{aligned}&|\partial_t g|>\epsilon_3>0, &&\forall|t|\le t_0,c_1\le\rho\le c_2;\\&\partial_{tt}g\le-\epsilon_4<0, &&\forall t_0\le|t|\le 1-\epsilon_0,c_1\le\rho\le c_2.\end{aligned}$$

The desired result then follows from Proposition 2.1. □

3.3 Strong convexity near the global minimizers $u=\pm e_1$: analysis of the limiting profile

In this section we shall show that in the small neighborhood of $u=\pm e_1$ where

$$||\hat{u}\cdot e_1|-1|\ll 1,\quad|\rho-1|\ll 1,$$

that the Hessian of the expectation of the loss function must be strictly positive definite. In yet other words $\mathbb{E}f$ must be strictly convex in this neighborhood so that $u=\pm e_1$ are the unique minimizers. To this end consider

$$h(u)=\frac{1}{2}(\mathbb{E}f-3)=\frac{1}{2}(1+2\beta)\rho^2-\mathbb{E}\sqrt{\beta\rho^2+\rho^2(a\cdot\hat{u})^2+(a\cdot e_1)^2}\sqrt{\beta\rho^2+2(a\cdot e_1)^2}, \quad (3.7)$$

where $a\sim\mathcal{N}(0,\mathrm{I}_n)$.

Theorem 3.5 (Strong convexity of $\mathbb{E}f$ when $\|u\pm e_1\|\ll 1$)**.** *Consider h defined by (3.7). There exist $0<\epsilon_0\ll 1$ and a positive constant γ_1 such that the following hold:*

1. *If $\|u-e_1\|_2\le\epsilon_0$, then for any $\xi\in\mathbb{S}^{n-1}$, we have*
$$\sum_{i,j=1}^n \xi_i\xi_j(\partial_i\partial_j h)(u)\ge\gamma_1>0.$$

2. *If $\|u+e_1\|_2\le\epsilon_0$, then for any $\xi\in\mathbb{S}^{n-1}$, we have*
$$\sum_{i,j=1}^n \xi_i\xi_j(\partial_i\partial_j h)(u)\ge\gamma_1>0.$$

Proof. We shall only consider the case $\|u-e_1\|_2\ll 1$. The other case $\|u+e_1\|_2\ll 1$ is similar and therefore omitted. Note that

$$\|u-e_1\|_2^2=\|\rho\hat{u}-e_1\|_2^2=(\rho-1)^2+2\rho(1-t)\le\epsilon_0^2,$$

where $t=\hat{u}\cdot e_1$. Thus for $0<\epsilon_0\ll 1$, we have

$$|\rho-1|\le\epsilon_0,\quad 1-\epsilon_0^2\le t\le 1.$$

We make a change of variable and write (recall $1-\epsilon_0^2\le\hat{u}\cdot e_1\to 1$),

$$\hat{u}=\sqrt{1-s^2}e_1+se^{\perp},\quad e^{\perp}\cdot e_1=0, e^{\perp}\in\mathbb{S}^{n-1},$$

where we assume $0\le s\ll 1$. Note that $s=\frac{|u'|}{\rho}$, and $u'=u-(u\cdot e_1)e_1=(0,u_2,\cdots,u_n)$. To calculate $\partial^2 h$ we need to compute the Hessian expressed in the (ρ,s) coordinate. It is not difficult to check that by (3.7), the value of $h(u)$ depends only on (ρ,s). Thus by a slight abuse of notation we write $h(u)=h(|u|,\frac{|u'|}{|u|})=h(\rho,s)$ (we denote $|u|=\|u\|_2$, $|u'|=\|u'\|_2$) and compute (below we assume $s>0$ so that $|u'|>0$)

$$\begin{aligned}\sum_{i,j=1}^n\xi_i\xi_j\partial_{ij}h=&\partial_{\rho\rho}h\cdot a^2+2a\partial_{\rho s}h\cdot\Big(-\frac{as}{\rho}+\frac{b}{\rho}\Big)+\partial_\rho h\cdot\Big(\frac{|\xi|^2-|\xi\cdot\hat{u}|^2}{\rho}\Big)+\partial_{ss}h\cdot\Big(\frac{a^2s^2-2abs}{\rho^2}\Big)\\&+\Big(\partial_{ss}h-\frac{1}{s}\partial_s h\Big)\frac{b^2}{\rho^2}+\partial_s h\cdot\Big(-\frac{|\xi|^2}{\rho^2}s+3\frac{a^2s}{\rho^2}-2\frac{ab}{\rho^2}+\frac{|\xi'|^2}{\rho^2 s}\Big).\end{aligned}$$

In the above computation, one does not need to worry about the formal singularity caused by $\frac{1}{s}$. Since (by Lemma B.6) $\partial_s h(\rho,s=0)=0$ for any $\rho>0$, we write

$$(\partial_s h)(\rho,s)\cdot\frac{1}{s}=\frac{(\partial_s h)(\rho,s)-(\partial_s h)(\rho,0)}{s}=\int_0^1(\partial_{ss}h)(\rho,\theta s)d\theta,\quad s>0.$$

In particular we have

$$\begin{aligned}&\lim_{s\to0^+}(\partial_s h)(\rho,s)\cdot\frac{1}{s}=(\partial_{ss}h)(\rho,0);\\&\Big|(\partial_{ss}h)(\rho,s)-\frac{1}{s}\partial_s h_1(\rho,s)\Big|=\mathcal{O}(s)\to0\quad\text{as}\quad s\to0.\end{aligned}$$

By using this observation and Lemma B.6, we obtain

$$\sum_{i,j=1}^n\xi_i\xi_j(\partial_{ij}h)(e_1)=(\partial_{\rho\rho}h)(1,0)\cdot a^2\Big|_{a=\xi_1}+(\partial_{ss}h)(1,0)\cdot|\xi'|^2\ge\gamma_0\cdot|\xi|^2,\quad\forall\xi\in\mathbb{S}^{n-1},$$

where $\gamma_0>0$ is a constant. Now for $\|u-e_1\|_2\ll1$, by using Lemma B.5 and Lemma B.6, we have

$$\begin{aligned}&\Big|\sum_{i,j=1}^n\xi_i\xi_j\Big((\partial_{ij}h)(e_1)-(\partial_{ij}h)(u)\Big)\Big|\\\lesssim&|(\partial_{\rho\rho}h)(\rho,s)-(\partial_{\rho\rho}h)(1,0)|+|(\partial_{\rho\rho}h)(\rho,s)|\cdot|(\xi\cdot\hat{u})^2-(\xi\cdot e_1)^2|\\&+|(\partial_{\rho s}h)(\rho,s)|\cdot(1+s)+|(\partial_\rho h)(\rho,s)|+|\partial_{ss}h(\rho,s)|\cdot(s+s^2)\\&+|\partial_s h(\rho,s)|+\Big|\frac{1}{\rho^2}\int_0^1(\partial_{ss}h)(\rho,\theta s)d\theta-(\partial_{ss}h)(1,0)\Big|\\&+\Big|(\partial_{ss}h)(\rho,s)-(\partial_s h)(\rho,s)\cdot\frac{1}{s}\Big|\cdot\frac{b^2}{\rho^2}\\\lesssim&\mathcal{O}(|\rho-1|+|s|+\|\hat{u}-e_1\|_2).\end{aligned}$$

It follows that if $\|u-e_1\|_2$ is sufficiently small, we then have

$$\sum_{i,j=1}^n\xi_i\xi_j(\partial_{ij}h)(u)\ge\frac{\gamma_0}{2}|\xi|^2,\quad\forall\xi\in\mathbb{S}^{n-1}.$$

This completes the proof. □

3.4 Near the global minimizer: strong convexity

In this section we show strong convexity of the loss function $f(u)$ near the global minimizer $u=\pm e_1$.

Theorem 3.6 (Strong convexity near the global minimizer)**.** *There exists $0<\epsilon_0\ll 1$ and a constant $\beta_1>0$ such that if $m\gtrsim n$, then the following hold with high probability:*

1. *If $\|u-e_1\|_2\le\epsilon_0$, then*

$$\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f)(u)\ge\gamma>0,\quad \forall\xi\in\mathbb{S}^{n-1},$$

   where γ is a constant.

2. *If $\|u+e_1\|_2\le\epsilon_0$, then*

$$\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f)(u)\ge\gamma>0,\quad \forall\xi\in\mathbb{S}^{n-1},$$

   where γ is a constant.

In yet other words, $f(u)$ is strongly convex in a sufficiently small neighborhood of $\pm e_1$.

Proof. Recall

$$f(u)=2f_0(u)+\frac{1}{m}\sum_{k=1}^{m}\Big((a_k\cdot u)^2+2\beta|u|^2+3(a_k\cdot e_1)^2\Big),$$

where

$$f_0(u)=-\frac{1}{m}\sum_{k=1}^{m}\sqrt{\beta|u|^2+(a_k\cdot u)^2+(a_k\cdot e_1)^2}\cdot\sqrt{\beta|u|^2+2(a_k\cdot e_1)^2}.$$

Clearly

$$\sum_{i,j=1}^{n}\xi_i\xi_j\partial_{ij}f(u)=2\Big(\frac{1}{m}\sum_{k=1}^{m}|a_k\cdot\xi|^2\Big)+4\beta|\xi|^2+2\sum_{i,j=1}^{n}\xi_i\xi_j\partial_{ij}f_0(u).$$

Obviously we have for $m \gtrsim n$, it holds with high probability that

$$\Big|\frac{1}{m}\sum_{k=1}^{m}|a_k\cdot\xi|^2-1\Big|\le\frac{\epsilon}{100},\quad \forall\xi\in\mathbb{S}^{n-1}.$$

By Lemma B.7, we have

$$\Big|\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f_0(u)-\mathbb{E}\partial_{ij}f_0(u))\Big|\le\frac{\epsilon}{100},\quad \forall\xi\in\mathbb{S}^{n-1},\quad \forall\frac{1}{3}\le\|u\|_2\le3.$$

Thus we have

$$\Big|\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f(u)-\mathbb{E}\partial_{ij}f(u))\Big|\le\frac{\epsilon}{100},\quad \forall\xi\in\mathbb{S}^{n-1},\quad \forall\frac{1}{3}\le\|u\|_2\le3.$$

The desired result then follows from Theorem 3.5 by taking $\epsilon>0$ sufficiently small. This completes the proof. □

We now complete the proof of the main theorem.

Proof of Theorem 3.1. We proceed in several steps.

1. By Theorem 3.2, we see that with high probability the function $f(u)$ has non-vanishing gradient in the regimes

$$0<\|u\|_2\le R_2=R_2(\beta)$$

and

$$\|u\|_2\ge R_1=R_1(\beta),$$

where $R_1>0$, $R_2>0$ depend only on β. Moreover the point $u=0$ is a local maximum point with strictly negative-definite Hessian.

2. By Theorem 3.6, there exists $\epsilon_0>0$ sufficiently small, such that with high probability, $f(u)$ is strongly convex in the neighborhood $\|u\pm e_1\|_2\le\epsilon_0$.

3. By Theorem 3.3, we have that with high probability

$$\|\nabla f\|_2>0,$$

if $|\rho-1|\ge c(\eta_0)$ and $||\hat{u}\cdot e_1|-1|\le\eta_0$. Here we recall $\rho=\|u\|_2$ and $u=\rho\hat{u}$. Observe that

$$\|u\pm e_1\|_2^2=(\rho-1)^2+2\rho(1\pm\hat{u}\cdot e_1).$$

By taking $\eta_0=\epsilon_0^2/100$, we see that $\|u\pm e_1\|_2>\epsilon_0$, $||\hat{u}\cdot e_1|-1|\le\eta_0$ must imply

$$|\rho-1|>\frac{\epsilon_0}{10}.$$

Thus it remains for us to treat the regime $\||\hat{u}\cdot e_1|-1|>\eta_0$, $R_1\le\|u\|_2\le R_2$.

4. In the regime $||\hat{u}\cdot e_1|-1|>\eta_0$, $\|u\|_2\sim 1$, we have by Theorem 3.4, with high probability it holds that either the function has a non-vanishing gradient at the point u, or the gradient vanishes at u, but f has a negative directional curvature at this point.

Thus we complete the proof. □

4 Numerical experiments

In this section, we demonstrate the numerical efficiency of our estimators by simple gradient descent and compare their performance with other competitive algorithms.

In a concurrent work [5], we considered the following piecewise Smoothed Amplitude loss (SAF):

$$\min_{u\in\mathbb{R}^n}\quad f(u)=\frac{1}{2m}\sum_{j=1}^m\left(\gamma\left(\frac{|a_j\cdot u|}{|a_j\cdot x|}\right)-1\right)^2\cdot|a_j\cdot x|^2$$

with the function $\gamma(t)$

$$\gamma(t):=\begin{cases}|t|, & |t|>\beta;\\ \dfrac{1}{2\beta}t^2+\dfrac{\beta}{2}, & |t|\le\beta.\end{cases}$$

In this work, our first Perturbed Amplitude Model (PAM1) is

$$\min_{u\in\mathbb{R}^n}\quad f(u)=\frac{1}{m}\sum_{j=1}^m\left(\sqrt{\beta|u|^2+(a_j\cdot u)^2}-\sqrt{\beta|u|^2+(a_j\cdot x)^2}\right)^2.$$

The second Perturbed Amplitude Model (PAM2) is

$$\min_{u\in\mathbb{R}^n}\quad f(u)=\frac{1}{m}\sum_{j=1}^m\left(\sqrt{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot x)^2}-\sqrt{\beta|u|^2+2(a_j\cdot x)^2}\right)^2.$$

We have show theoretically that any gradient descent algorithm will not get trapped in a local minimum for the loss functions above. Here we present numerical experiments to show that the estimators perform very well with randomized initial guess.

We use the following vanilla gradient descent algorithm

$$u_{k+1} = u_k - \mu \nabla f(u_k)$$

with a random initial guess to minimize the loss function $f(u)$ given above. The pseudocode for the algorithm is as follows.

Algorithm 4.1 Gradient descend algorithm based on our new models.

Input: Measurement vectors: $a_i \in \mathbb{R}^n, i = 1, \cdots, m$; Observations: $y \in \mathbb{R}^m$; Parameter β; Step size μ; Tolerance $\epsilon > 0$

1: Random initial guess $u_0 \in \mathbb{R}^n$.

2: For $k = 0, 1, 2, \cdots$, if $\|\nabla f(u_k)\| \geq \epsilon$ do

$$u_{k+1} = u_k - \mu \nabla f(u_k)$$

3: End for

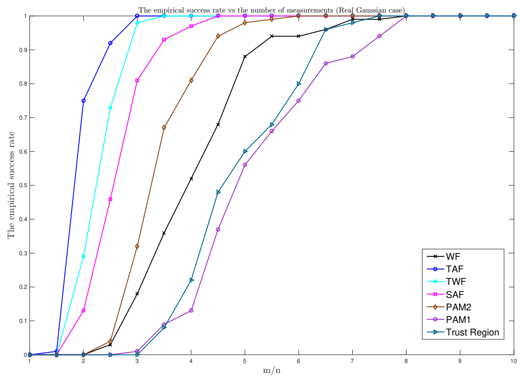
Output: The vector u_T.

The performance of our PAM1 and PAM2 algorithms are conducted via a series of numerical experiments in comparison against SAF, Trust Region [30], WF [3], TWF [6] and TAF [35]. Here, it is worth emphasizing that random initialization is used for SAF, Trust Region [30] and our PAM1, PAM2 algorithms while all other algorithms have adopted a spectral initialization. Our theoretical results are for real Gaussian case, but the algorithms can be easily adapted to the complex Gaussian and CDP cases. All experiments are carried out on a MacBook Pro with a 2.3GHz Intel Core i5 Processor and 8 GB 2133 MHz LPDDR3 memory.

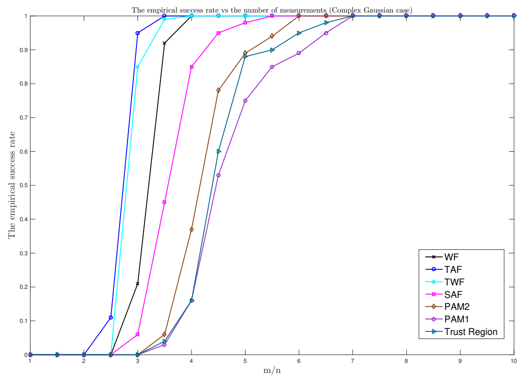
4.1 Recovery of 1D signals

In our numerical experiments, the target vector $x \in \mathbb{R}^n$ is chosen randomly from the standard Gaussian distribution and the measurement vectors a_i, $i = 1, \cdots, m$ are generated randomly from standard Gaussian distribution or CDP model. For the real Gaussian case, the signal $x \sim \mathcal{N}(0, I_n)$ and measurement vectors $a_i \sim \mathcal{N}(0, I_n)$ for $i = 1, \cdots, m$. For the complex Gaussian case, the signal $x \sim \mathcal{N}(0, I_n) + i\mathcal{N}(0, I_n)$ and measurement vectors $a_i \sim \mathcal{N}(0, I_n/2) + i\mathcal{N}(0, I_n/2)$. For the CDP model, we use masks of octanary patterns as in [3]. For simplicity, our parameters and step size are fixed for all experiments. Specifically, we adopt parameter $\beta = 1/2$ and step size $\mu = 1$ for SAF. We choose the parameter $\beta = 1$, step size $\mu = 0.6$ and $\mu = 2.5$ for PAM1 and PAM2, respectively. For Trust Region, WF, TWF and TAF, we use the code provided in the original papers with suggested parameters.

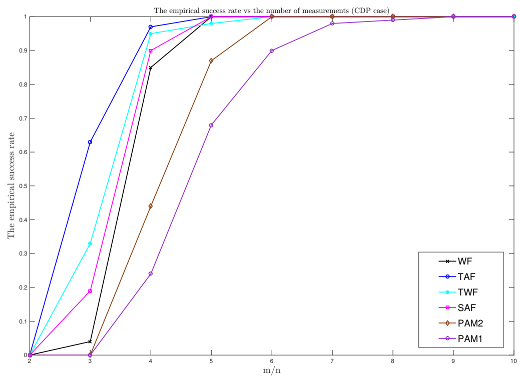
Example 4.1. In this example, we test the empirical success rate of PAM1, PAM2 versus the number of measurements. We conduct the experiments for the real Gaussian, complex Gaussian and CDP cases respectively. We choose $n=128$ and the maximum number of iterations is $T=2500$. For real and complex Gaussian cases, we vary m within the range $[n,10n]$. For CDP case, we set the ratio $m/n=L$ from 2 to 10. For each m, we run 100 times trials to calculate the success rate. Here, we say a trial to have successfully reconstructed the target signal if the relative error satisfies $\mathrm{dist}(u_T-x)/\|x\|\le 10^{-5}$. The results are plotted in Fig. 1. It can be seen that $6n$ Gaussian phaseless measurement or 7 octanary patterns are enough for exactly recovery for PAM2.



(a)



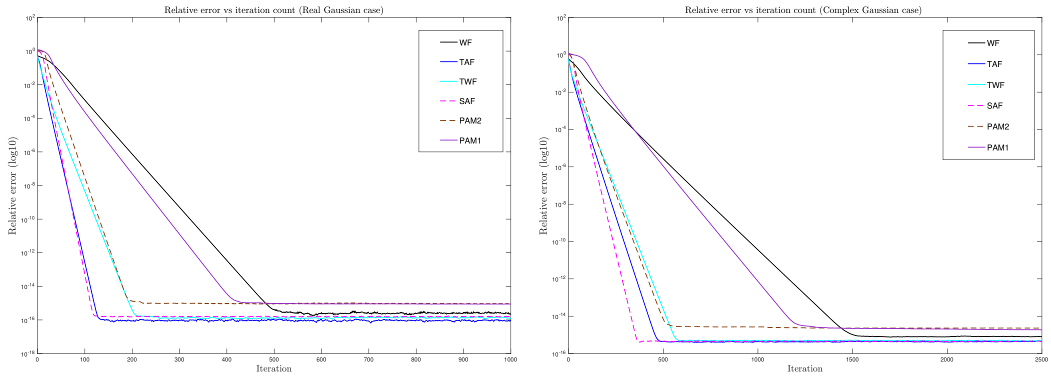
(b)



(c)

Figure 1: The empirical success rate for different m/n based on 100 random trails. (a) Success rate for real Gaussian case, (b) Success rate for complex Gaussian case, (c) Success rate for CDP case.

Example 4.2. In this example, we compare the convergence rate of PAM1, PAM2 with those of SAF, WF, TWF, TAF for real Gaussian and complex Gaussian cases.



(a) (b)

Figure 2: Relative error versus number of iterations for PAF, SAF, WF, TWF, and TAF method: (a) Real Gaussian case; (b) Complex Gaussian case.

We choose $n=128$ and $m=6n$. The results are presented in Fig. 2. Since PAM1 as well as PAM2 algorithm chooses a random initial guess according to the standard Gaussian distribution instead of adopting a spectral initialization, it sometimes need to escape the saddle points with a small number of iterations. Due to its high efficiency to escape the saddle points, it still performs well comparing with state-of-the-art algorithms with spectral initialization.

Example 4.3. In this example, we compare the time elapsed and the iteration needed for WF, TWF, TAF, SAF and our PAM1, PAM2 to achieve the relative error 10^{-5} and 10^{-10}, respectively. We choose $n=1000$ with $m=8n$. We adopt the same spectral initialization method for WF, TWF, TAF and the initial guess is obtained by power method with 50 iterations. We run 50 times trials to calculate the average time elapsed and iteration number for those algorithms. The results are shown in Table 1. The numerical results show that PAM2 takes around 15 and 42 iterations to escape the saddle points for the real and complex Gaussian cases, respectively.

4.2 Recovery of natural image

We next compare the performance of the above algorithms on recovering a natural image from masked Fourier intensity measurements. The image is the Milky Way Galaxy with resolution 1080×1920. The colored image has RGB channels. We use $L=20$ random octanary patterns to obtain the Fourier intensity measurements for each R/G/B channel as in [3]. Table 2 lists the averaged time elapsed and the

Table 1: Time Elapsed and Iteration Number among Algorithms on Gaussian Signals with $n=1000$.

Algorithm	Real Gaussian				Complex Gaussian			
	10^{-5}		10^{-10}		10^{-5}		10^{-10}	
	Iter	Time(s)	Iter	Time(s)	Iter	Time(s)	Iter	Time(s)
SAF	44	**0.1556**	68	**0.2276**	113	**1.3092**	190	**2.3596**
PAM1	108	3.3445	204	5.5768	291	35.8624	591	75.3231
PAM2	46	1.5816	84	2.1980	129	15.8295	239	27.6362
WF	125	4.4214	229	6.3176	304	34.6266	655	86.6993
TAF	29	0.2744	60	0.3515	100	1.7704	211	2.7852
TWF	40	0.3181	87	0.4274	112	1.9808	244	3.7432
Trust Region	**21**	2.9832	**29**	4.4683	**33**	19.1252	**42**	29.0338

Table 2: Time elapsed and iteration number among algorithms on recovery of galaxy image.

Algorithm	The Milky Way Galaxy			
	10^{-5}		10^{-10}	
	Iter	Time(s)	Iter	Time(s)
SAF	92	**202.47**	148	**351.21**
PAM1	198	462.27	306	710.27
PAM2	113	260.48	187	441.55
WF	158	381.7	277	621.63
TAF	**65**	223.89	**122**	368.22
TWF	68	315.14	145	566.84

iteration needed to achieve the relative error 10^{-5} and 10^{-10} over the three RGB channels. We can see that our algorithms have good performance comparing with state-of-the-art algorithms with spectral initialization. Furthermore, our algorithms perform well even with $L=10$ under 300 iterations, while WF fails. Fig. 3 shows the image recovered by PAM2.

4.3 Recovery of signals with noise

We now demonstrate the robustness of PAM1, PAM2 to noise and compare them with SAF, WF, TWF, TAF. We consider the noisy model $y_i=|\langle a_i,x\rangle|+\eta_i$ and add different level of Gaussian noises to explore the relationship between the signal-to-noise rate (SNR) of the measurements and the mean square error (MSE) of the recovered signal. Specifically, SNR and MSE are evaluated by

$$\mathrm{MSE}:=10\log_{10}\frac{\mathrm{dist}^2(u,x)}{\|x\|^2} \quad \text{and} \quad \mathrm{SNR}=10\log_{10}\frac{\sum_{i=1}^m|a_i^\top x|^2}{\|\eta\|^2},$$

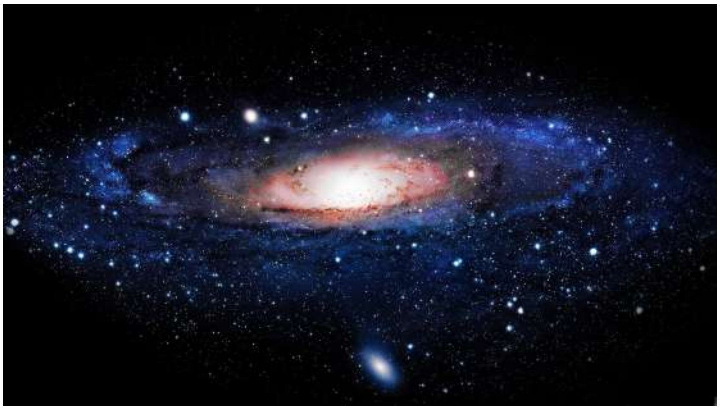
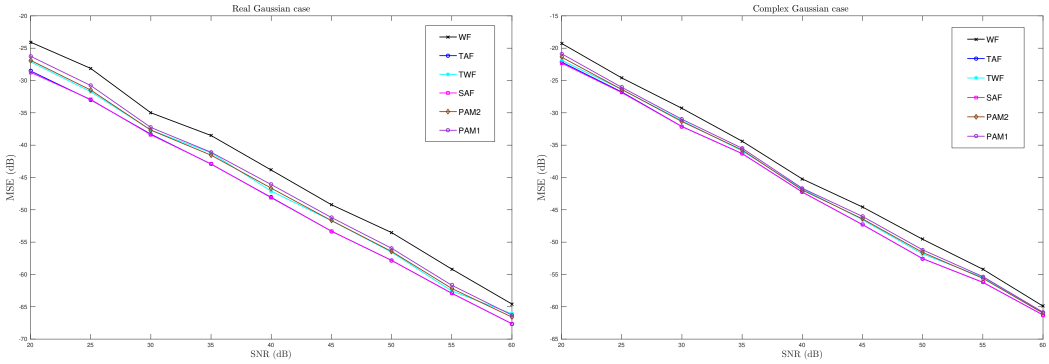


Figure 3: The milky way galaxy image: PAM2 with $L=10$ takes 300 iterations, computation time is 524.1s, relative error is 7.26×10^{-13}.



(a) (b)

Figure 4: SNR versus relative MSE on a dB-scale under the noisy Gaussian model: (a) Real Gaussian case; (b) Complex Gaussian case.

where u is the output of the algorithms given above after 2500 iterations. We choose $n=128$ and $m=8n$. The SNR varies from 20db to 60db. The result is shown in Fig. 4. We can see that our algorithms are stable for noisy phase retrieval.

Appendix A: auxiliary estimates for Section 2

Proof of Lemma 2.4. Recall that

$$\mathbb{E}\sqrt{\beta+X_t^2}\sqrt{\beta\rho^2+X_1^2}=:h_\infty(\rho,t).$$

Since $\rho\sim 1$ we shall slightly abuse notation and write $h_\infty(\rho,t)$ simply as $h(t)$ in this proof. Denote $g(x)=\sqrt{\beta+x^2}$. Clearly

$$g'=\frac{x}{\sqrt{\beta+x^2}},\quad g''=\beta(\beta+x^2)^{-\frac{3}{2}}.$$

Since $X_t=tX_1+\sqrt{1-t^2}Y_1$, where X_1 and Y_1 are independent standard 1D Gaussian random variables, we clearly have

$$h(t)=\int g(tx+\sqrt{1-t^2}y)k(x)\rho_1(x,y)dxdy,$$

where

$$k(x)=\sqrt{\beta\rho^2+x^2},\quad \rho_1(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.$$

Observe that

$$\begin{aligned}
&\partial_x(g(tx+\sqrt{1-t^2}y))=g'\cdot t;\\
&\partial_y(g(tx+\sqrt{1-t^2}y))=g'\cdot\sqrt{1-t^2};\\
&\partial_t(g(tx+\sqrt{1-t^2}y))=g'\cdot\Big(x-\frac{t}{\sqrt{1-t^2}}y\Big)=\frac{1}{\sqrt{1-t^2}}(x\partial_y g-y\partial_x g).
\end{aligned}$$

The third identity is the key to obtaining cancellation when calculating $h'(t)$ and $h''(t)$.

Observe that

$$(x\partial_y-y\partial_x)\rho_1\equiv 0,\quad \forall x,y\in\mathbb{R}.$$

Then clearly

$$\begin{aligned}
h'(t)&=\frac{1}{\sqrt{1-t^2}}\int\Big((x\partial_y-y\partial_x)(g(tx+\sqrt{1-t^2}y))\Big)k(x)\rho_1(x,y)dxdy\\
&=\frac{1}{\sqrt{1-t^2}}\int g(tx+\sqrt{1-t^2}y)\cdot yk'(x)\rho_1dxdy\\
&=\frac{1}{\sqrt{1-t^2}}2\int_{x>0,y>0}\Big(g(tx+\sqrt{1-t^2}y)-g(tx-\sqrt{1-t^2}y)\Big)\cdot yk'(x)\rho_1dxdy.
\end{aligned}$$

Clearly then

$$\begin{aligned}
&h'(t)>0,\quad 0<t<1;\\
&h'(t)<0,\quad -1<t<0.
\end{aligned}$$

Moreover,

$$|h'(t)| \le 4\int \|g'\|_\infty y^2 \cdot \rho_1 dxdy \lesssim 1, \quad \forall |t|<1.$$

Note that we can actually obtain $|h'(t)|\lesssim 1$ for all $|t|\le 1$. On the other hand, for $0<t<1$,

$$\begin{aligned} h'(t) &\gtrsim \int_{x>0,y>0} \frac{tyx}{\sqrt{\beta+(tx+\sqrt{1-t^2}y)^2}+\sqrt{\beta+(tx-\sqrt{1-t^2}y)^2}} \cdot y \cdot \frac{x}{\sqrt{\beta\rho^2+x}} \cdot e^{-\frac{x^2+y^2}{2}} dxdy \\ &\gtrsim \int_{x\sim 10,y\sim 10} \frac{tyx}{\sqrt{\beta+(tx+\sqrt{1-t^2}y)^2}+\sqrt{\beta+(tx-\sqrt{1-t^2}y)^2}} \cdot y \cdot \frac{x}{\sqrt{\beta\rho^2+x}} \cdot e^{-\frac{x^2+y^2}{2}} dxdy \\ &\gtrsim t. \end{aligned}$$

Note that the implied constants here are allowed to depend on β. Similarly one can show $-h'(t)\gtrsim |t|$ for $-1<t<0$. Next we treat $h''(t)$ in the regime $|t|\ll 1$. Observe that

$$\frac{d}{dt}\Big(\frac{1}{\sqrt{1-t^2}}\Big)=\mathcal{O}(t), \quad \frac{1}{\sqrt{1-t^2}}=1+\mathcal{O}(t^2).$$

Thus

$$\begin{aligned} h''(t) &= \mathcal{O}(t)+\frac{1}{\sqrt{1-t^2}}\int g'(tx+\sqrt{1-t^2}y)\Big(x-\frac{t}{\sqrt{1-t^2}}y\Big) yk'(x)\rho_1 dxdy \\ &= \mathcal{O}(t)+\int g'(tx+\sqrt{1-t^2}y)\Big(x-\frac{t}{\sqrt{1-t^2}}y\Big) yk'(x)\rho_1 dxdy. \end{aligned}$$

Observe that the contribution of $\frac{t}{\sqrt{1-t^2}}y$ is bounded by $\mathcal{O}(t)$. Then

$$h''(t)=\mathcal{O}(t)+\int g'(tx+\sqrt{1-t^2}y)y\frac{x^2}{\sqrt{\beta\rho^2+x^2}}\rho_1 dxdy.$$

Denote $x_t=tx+\sqrt{1-t^2}y$. Then $y=\frac{x_t-tx}{\sqrt{1-t^2}}$. The contribution due to $\frac{tx}{\sqrt{1-t^2}}$ is also $\mathcal{O}(t)$. Thus

$$h''(t)=\mathcal{O}(t)+\int g'(x_t)\frac{x_t}{\sqrt{1-t^2}}\frac{x^2}{\sqrt{\beta\rho^2+x^2}}\rho_1 dxdy.$$

Note that $g'(z)z=\frac{z^2}{\sqrt{\beta+z^2}}$. Thus for $|t|\ll 1$, if $9\le y\le 11$, $\frac{1}{4}\le x\le\frac{1}{2}$, then $x_t\sim 1$, and the main term is $\mathcal{O}(1)$. Thus

$$h''(t)\gtrsim 1$$

for all $|t|\ll 1$. $\square$

If we take the limit $\beta \to 0$ in

$$h(\rho,t)=\mathbb{E}\sqrt{\beta+X_t^2}\sqrt{\beta\rho^2+X_1^2}.$$

Then we obtain the expression

$$\mathbb{E}|X_t||X_1|.$$

Understanding this limiting case is of some importance for the case $\beta>0$. The following proposition gives a very explicit characterization.

Proposition A.1. *We have for $|t|<1$:*

$$\begin{aligned}
&\frac{\pi}{2}\mathbb{E}|X_t||X_1|=\Big(\frac{\pi}{2}-\arccos t\Big)t+\sqrt{1-t^2};\\
&\Big(\frac{\pi}{2}\mathbb{E}|X_t||X_1|\Big)'=\frac{\pi}{2}-\arccos t;\\
&\Big(\frac{\pi}{2}\mathbb{E}|X_t||X_1|\Big)''=\frac{1}{\sqrt{1-t^2}}.
\end{aligned}$$

Proof. We recall $X_t=tX+\sqrt{1-t^2}Y$, where $X=X_1\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$ are independent. Without loss of generality we can assume $0\le t<1$. Denote $t=\sin\theta_0$ where $0\le\theta_0<\frac{\pi}{2}$. Then by using polar coordinates, we have

$$\begin{aligned}
\frac{\pi}{2}\mathbb{E}|X_tX_1|&=\frac{1}{4}\int_{\mathbb{R}^2}|tx+\sqrt{1-t^2}y|\cdot|x|\cdot e^{-\frac{x^2+y^2}{2}}dxdy\\
&=\frac{1}{4}\int_0^{2\pi}|\sin(\theta+\theta_0)|\cdot|\cos\theta|d\theta\cdot\int_0^{\infty}r^3e^{-\frac{r^2}{2}}dr\\
&=\frac{1}{2}\int_0^{2\pi}|\sin(\theta+\theta_0)|\cdot|\cos\theta|d\theta\\
&=\int_0^{\pi}|\sin(\theta+\theta_0)|\cdot|\cos\theta|d\theta.
\end{aligned}$$

Now observe that

$$\begin{aligned}
&\int_0^{\pi}|\sin(\theta+\theta_0)|\cdot|\cos\theta|d\theta\\
=&\int_0^{\frac{\pi}{2}}\sin(\theta+\theta_0)\cdot\cos\theta d\theta-\int_{\frac{\pi}{2}}^{\pi-\theta_0}\sin(\theta+\theta_0)\cdot\cos\theta d\theta\\
&+\int_{\pi-\theta_0}^{\pi}\sin(\theta+\theta_0)\cdot\cos\theta d\theta.
\end{aligned}$$

The desired result then easily follows by an explicit computation. $\square$

Remark A.1. One may wonder why at $t=1$, the derivative is formally given by $\frac{\pi}{2}$ instead of being zero since $u=x$ should be a critical point. The reason is due to the artificial singularity introduced by our representation. To see this, one can consider the regular variable $t=\cos\theta$ with $\theta\in[0,\pi]$, then

$$f(\theta)=\Big(\frac{\pi}{2}-\theta\Big)\cos\theta+\sin\theta.$$

Then clearly $f'(0)=0$ and $f''(0)=-\frac{\pi}{2}<0$. On the other hand,

$$f(\theta)=\tilde{f}(t)=\tilde{f}(\cos\theta).$$

Then

$$f'(\theta)=\tilde{f}'(\cos\theta)(-\sin\theta).$$

Thus

$$\tilde{f}'(1)=\lim_{\theta\to 0}\frac{f'(\theta)}{-\sin\theta}=-f''(0)>0.$$

Lemma A.1. *Let X_i: $1\le i\le m$ be independent random variables with*

$$\max_{1\le i\le m}\mathbb{E}|X_i|^4\lesssim 1.$$

Then for any $t>0$,

$$\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{j=1}^m X_i-\text{mean}\Big|>t\Big)\lesssim\frac{1}{m^2t^4}.$$

Proof. Without loss of generality we can assume X_i has zero mean. The result then follows from the observation that

$$\mathbb{E}\Big(\sum_{j=1}^m X_j\Big)^4\lesssim\sum_{i<j}\mathbb{E}X_i^2X_j^2+\sum_i\mathbb{E}X_i^4\lesssim m^2.$$

Thus, we complete the proof. $\square$

Lemma A.2. *Let $\psi:\mathbb{R}\to\mathbb{R}$ be continuously differentiable such that*

$$\sup_{z\in\mathbb{R}}|\psi(z)|+\sup_{z\in\mathbb{R}}\sqrt{1+z^2}|\psi'(z)|\lesssim 1.$$

Let

$$\psi_0(R,z)=z\sqrt{R+z^2},\quad z\in\mathbb{R},\quad c_1\le R\le c_2,$$

where $0<c_1<c_2<\infty$ *are two fixed constants. For any* $0<\epsilon\le 1$*, if* $m\gtrsim n$*, then the following hold with high probability:*

$$\Big|\frac{1}{m}\sum_{j=1}^m \psi(a_j\cdot u)\psi_0(R,a_j\cdot e_1)-\mathrm{mean}\Big|\le\epsilon,\quad \forall u\in\mathbb{S}^{n-1},\quad \forall c_1\le R\le c_2.$$

Proof. Step 1. Let $0<\eta<\frac{1}{2}$ be a constant whose value will be chosen sufficiently small. Let $\phi\in C_c^\infty(\mathbb{R})$ be such that $0\le\phi(x)\le 1$ for all x, $\phi(x)=1$ for $|x|\le 1$, and $\phi(x)=0$ for $|x|\ge 2$. Denote

$$\langle x\rangle=\sqrt{1+x^2},\quad x\in\mathbb{R}.$$

Consider first the piece

$$\begin{aligned}
I_1=&\frac{1}{m}\sum_{j=1}^m\psi(a_j\cdot u)\phi\Big(\frac{a_j\cdot u}{\eta\langle a_j\cdot e_1\rangle}\Big)\psi_0(R,a_j\cdot e_1)\\
=&\frac{1}{m}\sum_{j=1}^m\psi(a_j\cdot u)\phi\Big(\frac{a_j\cdot u}{\eta\langle a_j\cdot e_1\rangle}\Big)\psi_0(R,a_j\cdot e_1)\phi\Big(\frac{a_j\cdot e_1}{K}\Big)\\
&+\frac{1}{m}\sum_{j=1}^m\psi(a_j\cdot u)\phi\Big(\frac{a_j\cdot u}{\eta\langle a_j\cdot e_1\rangle}\Big)\psi_0(R,a_j\cdot e_1)\Big(1-\phi\Big(\frac{a_j\cdot e_1}{K}\Big)\Big)\\
=&:I_{1,a}+I_{1,b},
\end{aligned}$$

where $K=\eta^{-\frac{1}{6}}$. Thanks to the cut-off $\phi(\frac{a_j\cdot e_1}{K})$, we have $|a_j\cdot e_1|\le 2K$ on its support. Thus

$$|\psi_0(R,a_j\cdot e_1)|\phi\Big(\frac{a_j\cdot e_1}{K}\Big)\lesssim K^2,$$

and

$$|I_{1,a}|\le\alpha_0K^2\frac{1}{m}\sum_{j=1}^m\phi\Big(\frac{a_j\cdot u}{\eta\langle 2K\rangle}\Big),$$

where $\alpha_0>0$ is an absolute constant. Clearly

$$\begin{aligned}
K^2\mathbb{E}\phi\Big(\frac{a_j\cdot u}{\eta\langle 2K\rangle}\Big)\le& K^2\frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}}\phi\Big(\frac{x}{\eta\langle 2K\rangle}\Big)dx\\
\lesssim&\eta^{-\frac{1}{3}}\cdot\eta^{\frac{5}{6}}=\eta^{\frac{1}{2}}.
\end{aligned}$$

By Bernstein's inequality, we have with high probability,

$$\Big|\frac{1}{m}\sum_{j=1}^{m}\phi\Big(\frac{a_j\cdot u}{\eta\langle 2K\rangle}\Big)-\text{mean}\Big|\le\eta.$$

Thus for $\eta>0$ sufficiently small,

$$|I_{1,a}|\lesssim\eta^{\frac{1}{2}}\le\frac{\epsilon}{10}.$$

For $I_{1,b}$, we have

$$|I_{1,b}|\le\alpha_1\cdot\frac{1}{m}\sum_{j=1}^{m}(a_j\cdot e_1)^2\Big(1-\phi\Big(\frac{a_j\cdot e_1}{K}\Big)\Big),$$

where $\alpha_1>0$ is a constant. Similar to the estimate in $I_{1,a}$, we have with high probability,

$$\alpha_1\cdot\frac{1}{m}\sum_{j=1}^{m}(a_j\cdot e_1)^2\Big(1-\phi\Big(\frac{a_j\cdot e_1}{K}\Big)\Big)\lesssim\eta\le\frac{\epsilon}{10}.$$

Thus with high probability, it holds that for sufficiently small η,

$$|I_1|\le\frac{\epsilon}{5},$$

By a simple estimate we have $|\mathbb{E}I_1|\le\frac{\epsilon}{5}$ for sufficiently small η. Thus

$$|I_1-\mathbb{E}I_1|\le\frac{2\epsilon}{5}.$$

Step 2. We now consider the main piece

$$I_2(u,R)=\frac{1}{m}\sum_{j=1}^{m}\psi(a_j\cdot u)\cdot\Big(1-\phi\Big(\frac{a_j\cdot u}{\eta\langle a_j\cdot e_1\rangle}\Big)\Big)\cdot\psi_0(R,a_j\cdot e_1).$$

Note that $\eta>0$ is fixed in Step 1. For simplicity we denote

$$X_j=a_j\cdot e_1,\quad h_1(z)=\psi(z)\cdot\Big(1-\phi\Big(\frac{z}{\eta\langle X_j\rangle}\Big)\Big).$$

Thanks to the cut-off $1-\phi(\frac{z}{\eta\langle X_j\rangle})$ and the fact that $|\psi'(z)|\lesssim\langle z\rangle^{-1}$, we have

$$|h_1'(z)|\lesssim\eta^{-1}\langle X_j\rangle^{-1}\lesssim\langle X_j\rangle^{-1},$$

where in the last inequality we have included η^{-1} into the implied constant. Since in this step $\eta>0$ is a fixed constant this will not cause any problem. Clearly then

$$|h_1(z)-h_1(\tilde{z})|\lesssim\langle X_j\rangle^{-1}|z-\tilde{z}|,\quad \forall z,\tilde{z}\in\mathbb{R}. \tag{A.1}$$

Also

$$|\psi_0(R,X_j)-\psi_0(\tilde{R},X_j)|\le|X_j|\cdot|R-\tilde{R}|^{\frac{1}{2}},\quad \forall c_1\le R,\quad \tilde{R}\le c_2. \tag{A.2}$$

We shall need these important estimates below.

Let $\delta>0$ be a small constant whose smallness will be specified later. We choose a δ-net F_δ covering the set $\mathbb{S}^{n-1}\times\{R: c_1\le R\le c_2\}$. We endow the set $S^{n-1}\times\{R: c_1\le R\le c_2\}$ with the simple metric:

$$d((u,R),(\tilde{u},\tilde{R}))=\|u-\tilde{u}\|_2+|R-\tilde{R}|.$$

Note that

$$\mathrm{Card}(F_\delta)\le\exp(C_\delta n),$$

where $C_\delta>0$ depends only on δ. By Bernstein's inequality, we have for any $0<\eta_1\le\frac{1}{2}$,

$$\mathbb{P}\Big(\sup_{(u,R)\in F_\delta}\Big|I_2(u,R)-\mathrm{mean}\Big|\ge\eta_1\Big)\le 2e^{C_\delta n}\cdot e^{-c\eta_1^2 m}.$$

Thus with high probability and taking $\eta_1=\frac{\epsilon}{10}$, we have

$$|I_2(u,R)-\mathbb{E}I_2(u,R)|\le\frac{\epsilon}{10},\quad \forall (u,R)\in F_\delta.$$

Now let $(u,R)\in F_\delta$, and consider any $(\tilde{u},\tilde{R})$ such that

$$\|u-\tilde{u}\|_2+|R-\tilde{R}|\le\delta.$$

By using the estimates (A.1), (A.2), we have

$$\begin{aligned}
&|I_2(u,R)-I_2(\tilde{u},\tilde{R})|\\
=&\Big|\frac{1}{m}\sum_{j=1}^m(h_1(a_j\cdot u)\psi_0(R,X_j)-h_1(a_j\cdot\tilde{u})\psi_0(\tilde{R},X_j))\Big|\\
\le&\frac{1}{m}\sum_{j=1}^m|h_1(a_j\cdot u)-h_1(a_j\cdot\tilde{u})||\psi_0(R,X_j)|+\frac{1}{m}\sum_{j=1}^m|h_1(a_j\cdot\tilde{u})||\psi_0(R,X_j)-\psi_0(\tilde{R},X_j)|\\
\le&\alpha_2\frac{1}{m}\sum_{j=1}^m\langle X_j\rangle^{-1}|a_j\cdot(u-\tilde{u})||X_j|\cdot\langle X_j\rangle+\alpha_2\frac{1}{m}\sum_{j=1}^m|R-\tilde{R}|^{\frac{1}{2}}|X_j|\langle X_j\rangle\\
\le&\alpha_2\frac{1}{m}\sum_{j=1}^m\Big(\frac{1}{2}|a_j\cdot(u-\tilde{u})|^2\cdot\delta^{-1}+\frac{1}{2}\delta|X_j|^2\Big)+\alpha_2|R-\tilde{R}|^{\frac{1}{2}}\frac{1}{m}\sum_{j=1}^m(|X_j|^2+1),
\end{aligned}$$

where $\alpha_2>0$ is an absolute constant. By Bernstein's inequality, it holds with high probability that

$$\frac{1}{m}\sum_{j=1}^{m}|a_j\cdot v|^2\leq 2,\quad \forall v\in\mathbb{S}^{n-1}.$$

Thus

$$|I_2(u,R)-I_2(\tilde{u},\tilde{R})|\leq 10\alpha_2(\delta+\delta^{\frac{1}{2}}).$$

Also it is easy to check that

$$|\mathbb{E}(I_2(u,R)-I_2(\tilde{u},\tilde{R}))|\leq 10\alpha_2(\delta+\delta^{\frac{1}{2}}).$$

Therefore

$$|I_2(u,R)-\mathbb{E}I_2(u,R)-(I_2(\tilde{u},\tilde{R}),-\mathbb{E}I_2(\tilde{u},\tilde{R})|\leq 20\alpha_2(\delta+\delta^{\frac{1}{2}}).$$

Now take δ such that

$$20\alpha_2(\delta+\delta^{\frac{1}{2}})\leq\frac{\epsilon}{10}.$$

We then obtain (with high probability)

$$|I_2(u,R)-\mathbb{E}I_2(u,R)|\leq\frac{\epsilon}{5},\quad \forall u\in S^{n-1},\quad \forall c_1\leq R\leq c_2.$$

Together with the estimate of I_1 in Step 1, we obtain the desired conclusion. $\square$

Lemma A.3. *Let $\psi:\mathbb{R}\to\mathbb{R}$ be Lipschitz continuous such that*

$$\sup_{z\in\mathbb{R}}\frac{|\psi(z)|}{1+|z|}+\sup_{z\neq\tilde{z}\in\mathbb{R}}\frac{|\psi(z)-\psi(\tilde{z})|}{|z-\tilde{z}|}\lesssim 1.$$

Let $0<c_1<c_2<\infty$ be two fixed constants. For any $0<\epsilon\leq 1$, if $m\gtrsim n$, then the following hold with high probability:

$$\Big|\frac{1}{m}\sum_{j=1}^{m}\psi(a_j\cdot u)\sqrt{R+(a_j\cdot e_1)^2}-\text{mean}\Big|\leq\epsilon,\quad \forall u\in\mathbb{S}^{n-1},\quad \forall c_1\leq R\leq c_2.$$

Proof. The main point is use a δ-covering of the set $\mathbb{S}^{n-1}\times\{R:c_1\leq R\leq c_2\}$. Note that

$$\begin{aligned}&\Big|\psi(a_j\cdot u)\sqrt{R+(a_j\cdot e_1)^2}-\psi(a_j\cdot\tilde{u})\sqrt{\tilde{R}+(a_j\cdot e_1)^2}\Big|\\ \leq&|\psi(a_j\cdot u)-\psi(a_j\cdot\tilde{u})|\sqrt{R+(a_j\cdot e_1)^2}+|\psi(a_j\cdot\tilde{u})|\cdot\Big|\sqrt{R+(a_j\cdot e_1)^2}-\sqrt{\tilde{R}+(a_j\cdot e_1)^2}\Big|\\ \lesssim&|a_j\cdot(u-\tilde{u})|(1+|a_j\cdot e_1|)+(1+|a_j\cdot\tilde{u}|)\cdot|R-\tilde{R}|^{\frac{1}{2}}.\end{aligned}$$

The argument is then similar to that in Lemma A.2. We omit details. $\square$

Consider

$$h=\frac{1}{m}\sum_{j=1}^{m}\sqrt{\beta+(a_j\cdot\hat{u})^2}\cdot\sqrt{R+X_j^2},$$

where

$$\begin{aligned}&\hat{u}=te_1+\sqrt{1-t^2}e^{\perp},\quad |t|<1,\quad e^{\perp}\cdot e_1=0,\quad e^{\perp}\in\mathbb{S}^{n-1};\\ &X_j=a_j\cdot e_1,\quad 0<c_1\le R\le c_2<\infty.\end{aligned}$$

In the above we take $c_1>0$, $c_2>0$ as two fixed constants. In our original model, $R=\beta\rho^2$ and $\rho\sim 1$, and therefore this assumption is quite natural. In the lemma below we shall study h in the regime

$$|t|\le 1-\epsilon_0,$$

where $0<\epsilon_0\ll 1$. The smallness of ϵ_0 will be needed later when we study the regime $||\hat{u}\cdot e_1|-1|\ll 1$. Here we shall show that away from $|t|=1$ we have good control of h.

Lemma A.4. *Let $0<\epsilon_0\ll 1$ be fixed. For any $0<\epsilon\le 1$, if $m\gtrsim n$, then with high probability it holds that*

$$|\partial_t h-\mathbb{E}\partial_t h|\le\epsilon,\quad \forall |t|\le 1-\epsilon_0,\quad e^{\perp}\cdot e_1=0,\quad e^{\perp}\in\mathbb{S}^{n-1},\quad c_1\le R\le c_2.$$

Proof. Denote $g(x)=\sqrt{\beta+x^2}$ and

$$Z_j=a_j\cdot\hat{u}=tX_j+\sqrt{1-t^2}Y_j,\quad Y_j=a_j\cdot e^{\perp}.$$

Clearly

$$\begin{aligned}\frac{d}{dt}Z_j&=X_j-\frac{t}{\sqrt{1-t^2}}Y_j\\ &=\frac{1}{1-t^2}X_j-\frac{t}{1-t^2}Z_j.\end{aligned}$$

Therefore

$$\begin{aligned}\partial_t h&=\frac{1}{1-t^2}\cdot\frac{1}{m}\sum_{j=1}^{m}g'(Z_j)X_j\sqrt{R+X_j^2}-\frac{t}{1-t^2}\cdot\frac{1}{m}\sum_{j=1}^{m}g'(Z_j)Z_j\sqrt{R+X_j^2}\\ &=:\frac{1}{1-t^2}H_1-\frac{t}{1-t^2}H_2.\end{aligned}$$

By Lemma A.2, it holds with high probability that

$$|H_1-\mathbb{E}H_1|\le(1-\epsilon_0^2)\cdot\frac{\epsilon}{3},\quad \forall\hat{u}\in\mathbb{S}^{n-1},\quad c_1\le R\le c_2.$$

For H_2, we observe that

$$g'(x)x=\frac{x^2}{\sqrt{\beta+x^2}}.$$

By Lemma A.3, it then holds with high probability that

$$|H_2-\mathbb{E}H_2|\le(1-\epsilon_0^2)\cdot\frac{\epsilon}{3},\quad \forall\hat{u}\in\mathbb{S}^{n-1},\quad c_1\le R\le c_2.$$

The desired result then easily follows. $\square$

Lemma A.5. *Let $0<\epsilon_0\ll 1$ be fixed. For any $0<\epsilon\le\frac{1}{2}$, if $m\gtrsim n$, then with high probability it holds that*

$$\partial_{tt}h\ge\mathbb{E}\partial_{tt}h-\epsilon,\quad \forall|t|\le 1-\epsilon_0,\quad e^\perp\cdot e_1=0,\quad e^\perp\in\mathbb{S}^{n-1},\quad c_1\le R\le c_2.$$

Furthermore, it holds with probability at least $1-\mathcal{O}(m^{-2})$ that

$$|\partial_{tt}h-\mathbb{E}\partial_{tt}h|\le\epsilon,\quad \forall|t|\le 1-\epsilon_0,\quad e^\perp\cdot e_1=0,\quad e^\perp\in\mathbb{S}^{n-1},\quad c_1\le R\le c_2.$$

Proof. We adopt the same notation as in Lemma A.4. Observe that

$$\begin{aligned}\partial_{tt}h&=\frac{1}{m}\sum_{j=1}^m g'(Z_j)\frac{d^2}{dt^2}Z_j\sqrt{R+X_j^2}+\frac{1}{m}\sum_{j=1}^m g''(Z_j)\Big(\frac{d}{dt}Z_j\Big)^2\sqrt{R+X_j^2}\\&=:H_1+H_2.\end{aligned}$$

We first deal with H_1. Note that

$$\frac{d^2}{dt^2}Z_j=-(1-t^2)^{-\frac{3}{2}}Y_j.$$

Since

$$Y_j=\frac{1}{\sqrt{1-t^2}}(Z_j-tX_j),$$

we obtain

$$H_1=-(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m g'(Z_j)Z_j\sqrt{R+X_j^2}+\frac{t}{(1-t^2)^2}\frac{1}{m}\sum_{j=1}^m g'(Z_j)X_j\sqrt{R+X_j^2}.$$

By similar estimates as in Lemma A.4, we have with high probability,

$$\Big|\frac{1}{m}\sum_{j=1}^{m} g'(Z_j)Z_j\sqrt{R+X_j^2}-\text{mean}\Big| \le \frac{\epsilon}{20}\cdot(1-\epsilon_0^2)^2, \qquad \forall u\in\mathbb{S}^{n-1}, \quad c_1\le R\le c_2;$$

$$\Big|\frac{1}{m}\sum_{j=1}^{m} g'(Z_j)X_j\sqrt{R+X_j^2}-\text{mean}\Big| \le \frac{\epsilon}{20}\cdot(1-\epsilon_0^2)^2, \qquad \forall u\in\mathbb{S}^{n-1}, \quad c_1\le R\le c_2.$$

Thus

$$|H_1-\mathbb{E}H_1|\le\frac{\epsilon}{10}, \quad \forall |t|\le 1-\epsilon_0, \quad e^\perp\cdot e_1=0, \quad e^\perp\in\mathbb{S}^{n-1}, \quad c_1\le R\le c_2.$$

Next we deal with H_2. Observe that

$$g''(x)=\beta(\beta+x^2)^{-\frac{3}{2}}>0.$$

Let $\phi\in C_c^\infty(\mathbb{R})$ be such that $0\le\phi(x)\le 1$ for all x, $\phi(x)=1$ for $|x|\le 1$ and $\phi(x)=0$ for $|x|\ge 2$. Then

$$H_2=H_3+\frac{1}{m}\sum_{j=1}^{m} g''(Z_j)\cdot\Big(\frac{d}{dt}Z_j\Big)^2\cdot\Big(1-\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\Big)\sqrt{R+X_j^2},$$

where $H_3\ge 0$ is given by

$$H_3=\frac{1}{m}\sum_{j=1}^{m} g''(Z_j)\cdot\Big(\frac{d}{dt}Z_j\Big)^2\cdot\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\sqrt{R+X_j^2}.$$

We first show that if $0<\eta\le\frac{1}{2}$ is taken sufficiently small, then

$$\mathbb{E}H_3\le\frac{\epsilon}{20}; \tag{A.3}$$

and with probability at least $1-\mathcal{O}(m^{-2})$,

$$H_3\le\frac{\epsilon}{20}, \quad \forall|t|\le 1-\epsilon_0, \quad e^\perp\cdot e_1=0, \quad e^\perp\in\mathbb{S}^{n-1}, \quad c_1\le R\le c_2. \tag{A.4}$$

Here we stress that since $H_3\ge 0$, if we only care about the lower bound, we can just discard it in order to obtain a high-in-probability statement. On the other hand, to get a two-way bound of H_3, we need to work with weaker statements due to the high-moment terms (i.e., more than quadratic) of X_j in H_3.

Recall that

$$\begin{aligned}\Big|\frac{d}{dt}Z_j\Big| &= \Big|X_j - \frac{t}{\sqrt{1-t^2}}Y_j\Big| = \Big|\frac{1}{1-t^2}X_j - \frac{t}{1-t^2}Z_j\Big| \\ &\lesssim |X_j|+|Z_j|.\end{aligned}$$

We have

$$\begin{aligned}H_3 &\lesssim \frac{1}{m}\sum_{j=1}^m \phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\langle Z_j\rangle^{-3}\cdot(X_j^2+Z_j^2)\cdot(1+|X_j|) \\ &\lesssim \frac{1}{m}\sum_{j=1}^m \phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)(1+|X_j|^3).\end{aligned}$$

Let $K=\eta^{-\frac{1}{8}}$. Then

$$H_3 \lesssim K^3\frac{1}{m}\sum_{j=1}^m \phi\Big(\frac{Z_j}{\eta\langle 2K\rangle}\Big)+\frac{1}{m}\sum_{j=1}^m(1+|X_j|^3)\Big(1-\phi\Big(\frac{X_j}{K}\Big)\Big).$$

Clearly then for $\eta>0$ sufficiently small,

$$\begin{aligned}\mathbb{E}H_3 &\lesssim K^3\mathbb{E}\phi\Big(\frac{Z_1}{\eta\langle 2K\rangle}\Big)+\mathbb{E}(1+|X_1|^3)\Big(1-\phi\Big(\frac{X_1}{K}\Big)\Big) \\ &\le \frac{\epsilon}{20}.\end{aligned}$$

For $m\gtrsim n$, it holds with high probability that

$$\Big|\frac{1}{m}\sum_{j=1}^m \phi\Big(\frac{Z_j}{\eta\langle 2K\rangle}\Big)-\text{mean}\Big|\le\eta.$$

On the other hand, by Lemma A.1, it holds with probability at least $1-\mathcal{O}(m^{-2})$ that

$$\Big|\frac{1}{m}\sum_{j=1}^m(1+|X_j|^3)\Big(1-\phi\Big(\frac{X_j}{K}\Big)\Big)-\text{mean}\Big|\le\eta, \quad \forall u\in\mathbb{S}^{n-1}.$$

Thus for $\eta>0$ sufficiently small, (A.3) and (A.4) hold. Now we consider the main piece

$$H_4=\frac{1}{m}\sum_{j=1}^m g''(Z_j)\cdot\Big(\frac{d}{dt}Z_j\Big)^2\cdot\Big(1-\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\Big)\sqrt{R+X_j^2}.$$

By using

$$\begin{aligned}\Big(\frac{d}{dt}Z_j\Big)^2 &= \Big(\frac{1}{1-t^2}X_j - \frac{t}{1-t^2}Z_j\Big)^2 \\ &= \frac{1}{(1-t^2)^2}(X_j^2 + t^2 Z_j^2 - 2tX_jZ_j).\end{aligned}$$

Then

$$\begin{aligned}H_4 =& (1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m g''(Z_j)\Big(1-\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\Big)X_j^2\sqrt{R+X_j^2} \\ &+ (1-t^2)^{-2}t^2\frac{1}{m}\sum_{j=1}^m g''(Z_j)Z_j^2\Big(1-\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\Big)\sqrt{R+X_j^2} \\ &- 2t(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m g''(Z_j)Z_j\Big(1-\phi\Big(\frac{Z_j}{\eta\langle X_j\rangle}\Big)\Big)X_j\sqrt{R+X_j^2}.\end{aligned}$$

Define

$$h_j(x) = g''(x)\cdot\Big(1-\phi\Big(\frac{x}{\eta\langle X_j\rangle}\Big)\Big).$$

Clearly, thanks to the cut-off $1-\phi$, we have

$$\begin{aligned}&\|h_j\|_\infty \lesssim \langle X_j\rangle^{-3}; \\ &\|h_j'\|_\infty \lesssim \langle X_j\rangle^{-4}.\end{aligned}$$

It is then easy to check that the summands in H_4 are bounded. Moreover

$$|h_j(a_j\cdot u) - h_j(a_j\cdot\tilde{u})| \lesssim |a_j\cdot(u-\tilde{u})|\cdot\langle X_j\rangle^{-4}.$$

Similar bounds also hold for the other summands in H_4. Thus by a similar union bound argument as in Lemma A.2 (and taking care of the covering in the R-variable), we have with high probability that

$$|H_4 - \mathbb{E}H_4| \le \frac{\epsilon}{20}, \quad \forall |t| \le 1-\epsilon_0, \quad e^\perp\cdot e_1 = 0, \quad e^\perp \in \mathbb{S}^{n-1}, \quad c_1 \le R \le c_2.$$

Collecting all the estimates, we then obtain the desired estimate for $\partial_{tt}h$. $\square$

Lemma A.6. *Let $X\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$ be independent. Define*

$$\begin{aligned}&H(\rho,s) = \mathbb{E}\sqrt{\beta + (\sqrt{1-s^2}X + sY)^2}\sqrt{\beta\rho^2 + X^2}; \\ &h(\rho,s) = \frac{1}{2}(1+2\beta)\rho^2 - \rho H(\rho,s).\end{aligned}$$

Then it holds that

$$\sup_{|\rho-1|\ll 1,|s|\ll 1}\sum_{j=1}^{3}(|\partial^j H|+|\partial^j h|)\lesssim 1,$$

where $\partial=\partial_\rho$ *or* ∂_s*.*

Proof. Clearly it suffices for us to prove the estimate for H since the estimate for h will follow from it. We first deal with $\partial_{sss}H$ which appears to be the most difficult case and simultaneously $\partial_s H$, $\partial_{ss}H$. In some terms we shall even exhibit $(\beta,\ \rho)$-independent bounds which will be of interest for future investigations. Denote $A=\sqrt{1-s^2}x+sy$. Then

$$\begin{aligned}
&\partial_s A=-\frac{s}{\sqrt{1-s^2}}x+y;\\
&\partial_y A=s,\quad \partial_x A=\sqrt{1-s^2};\\
&\partial_s A=\frac{1}{\sqrt{1-s^2}}(-x\partial_y+y\partial_x)A.
\end{aligned}$$

Now we have

$$2\pi H=\int\sqrt{\beta+A^2}\sqrt{\beta\rho^2+x^2}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Since

$$\begin{aligned}
&\partial_s(\sqrt{\beta+A^2})=\frac{1}{\sqrt{1-s^2}}(-x\partial_y+y\partial_x)(\sqrt{\beta+A^2}),\\
&(-x\partial_y+y\partial_x)(e^{-\frac{x^2+y^2}{2}})=0,
\end{aligned}$$

we obtain (by using integration by parts) that

$$\begin{aligned}
2\pi\partial_s H&=\frac{1}{\sqrt{1-s^2}}\int\sqrt{\beta+A^2}(-y\partial_x)(\sqrt{\beta\rho^2+x^2})e^{-\frac{x^2+y^2}{2}}dxdy\\
&=-\frac{1}{\sqrt{1-s^2}}\int\sqrt{\beta+A^2}\frac{xy}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy.
\end{aligned}$$

Note that the pre-factor $\frac{1}{\sqrt{1-s^2}}$ is smooth in the regime $|s|\ll 1$, therefore to compute the higher order ∂_s-derivatives of H, it suffices for us to treat

$$H_1=\int\sqrt{\beta+A^2}\frac{xy}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Then

$$\partial_s H_1 = \frac{1}{\sqrt{1-s^2}} \int \sqrt{\beta+A^2}(x\partial_y - y\partial_x)\Big(\frac{xy}{\sqrt{\beta\rho^2+x^2}}\Big) e^{-\frac{x^2+y^2}{2}} dxdy.$$

The most difficult term is the piece corresponding to $y\partial_x$. Thus we consider

$$\begin{aligned} H_2 &= \int \sqrt{\beta+A^2}(y\partial_x)\Big(\frac{xy}{\sqrt{\beta\rho^2+x^2}}\Big) e^{-\frac{x^2+y^2}{2}} dxdy \\ &= \int \sqrt{\beta+A^2}\cdot y^2 \cdot \frac{\beta\rho^2}{(\beta\rho^2+x^2)^{\frac{3}{2}}} e^{-\frac{x^2+y^2}{2}} dxdy. \end{aligned}$$

Thus

$$\partial_s H_2 = \int \frac{A}{\sqrt{\beta+A^2}} \Big(-\frac{s}{\sqrt{1-s^2}} x + y\Big) y^2 \cdot \frac{\beta\rho^2}{(\beta\rho^2+x^2)^{\frac{3}{2}}} e^{-\frac{x^2+y^2}{2}} dxdy.$$

The piece corresponding to $-\frac{s}{\sqrt{1-s^2}}x$ is clearly fine. So we only need to treat

$$H_3 = \int \frac{A}{\sqrt{\beta+A^2}} y^3 \cdot \frac{\beta\rho^2}{(\beta\rho^2+x^2)^{\frac{3}{2}}} e^{-\frac{x^2+y^2}{2}} dxdy.$$

Observe that for $\eta > 0$,

$$\int_{\mathbb{R}} \frac{\eta^2}{(\eta^2+x^2)^{\frac{3}{2}}} dx = \int_{\mathbb{R}} \frac{1}{(1+x^2)^{\frac{3}{2}}} dx.$$

Thus H_3 is bounded by an absolute constant. Collecting the estimates, we have

$$|\partial_{sss} H| \lesssim 1.$$

Now we deal with $\partial_\rho H$, $\partial_{\rho\rho} H$, and $\partial_{\rho\rho\rho} H$. This case is easy. Denote $B = \beta\rho^2 + x^2$. Then

$$\begin{aligned} &\partial_\rho(\sqrt{B}) = B^{-\frac{1}{2}}\beta\rho; \\ &\partial_{\rho\rho}(\sqrt{B}) = B^{-\frac{1}{2}}\beta - B^{-\frac{3}{2}}\beta^2\rho^2; \\ &\partial_{\rho\rho\rho}(\sqrt{B}) = -B^{-\frac{3}{2}}\beta^2\rho + 3B^{-\frac{5}{2}}(\beta\rho)^3 - B^{-\frac{3}{2}}\beta^2 2\rho. \end{aligned}$$

Clearly all terms are bounded and we have

$$|\partial_\rho H| + |\partial_{\rho\rho} H| + |\partial_{\rho\rho\rho} H| \lesssim 1.$$

Next clearly $\partial_{\rho s}H$ and $\partial_{\rho\rho s}H$ are OK. We only need to treat $\partial_{\rho ss}H$. The main term of $\partial_{ss}H$ is

$$H_4=\int\frac{A}{\sqrt{\beta+A^2}}\Big(-\frac{s}{\sqrt{1-s^2}}x+y\Big)\frac{xy}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Now

$$\partial_\rho H_4=-\int\frac{A}{\sqrt{\beta+A^2}}\Big(-\frac{s}{\sqrt{1-s^2}}x+y\Big)\frac{xy\beta\rho}{(\beta\rho^2+x^2)^{\frac{3}{2}}}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Clearly for any $0<\eta\lesssim 1$,

$$\int\frac{\eta^2|x|}{(\eta^2+x^2)^{\frac{3}{2}}}dx=\eta\int\frac{|x|}{(1+x^2)^{\frac{3}{2}}}dx<\infty.$$

Thus $\partial_{\rho ss}H$ is also OK for us. $\square$

Lemma A.7 (Calculation of $\partial^2 h$ at $(\rho=1,\ s=0)$)**.** *Let*

$$\begin{aligned}
&H(\rho,s)=\mathbb{E}\sqrt{\beta+(\sqrt{1-s^2}X+sY)^2}\sqrt{\beta\rho^2+X^2};\\
&h(\rho,s)=\frac{1}{2}(1+2\beta)\rho^2-\rho H(\rho,s).
\end{aligned}$$

Then at $\rho=1$, $s=0$, we have

$$\begin{aligned}
&(\partial_{\rho\rho}H)(1,0)=-\gamma_1<0, \quad (\partial_{\rho s}H)(1,0)=0;\\
&(\partial_{ss}H)(1,0)=-\gamma_2<0;\\
&(\partial_s h)(\rho,0)=0, \quad \forall\rho>0, \quad (\partial_\rho h)(1,0)=0;\\
&(\partial_{\rho\rho}h)(1,0)=\gamma_3>0, \quad (\partial_{\rho s}h)(1,0)=0;\\
&(\partial_{ss}h)(1,0)=\gamma_4>0,
\end{aligned}$$

where $\gamma_i>0$, $i=1,\cdots,4$ are constants depending on β.

Proof. Calculation of $\partial_{ss}H$. Denote $A=\sqrt{1-s^2}x+sy$. Then

$$\begin{aligned}
&\partial_s A=-\frac{s}{\sqrt{1-s^2}}x+y, \quad \partial_y A=s, \quad \partial_x A=\sqrt{1-s^2};\\
&\partial_s A=\frac{1}{\sqrt{1-s^2}}(-x\partial_y+y\partial_x)A.
\end{aligned}$$

Now we have

$$2\pi H=\int\sqrt{\beta+A^2}\sqrt{\beta\rho^2+x^2}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Then

$$\begin{aligned}2\pi\partial_s H&=\frac{1}{\sqrt{1-s^2}}\int\sqrt{\beta+A^2}(-y\partial_x)(\sqrt{\beta\rho^2+x^2})e^{-\frac{x^2+y^2}{2}}dxdy\\&=-\frac{1}{\sqrt{1-s^2}}\int\sqrt{\beta+A^2}\frac{xy}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

One should observe that $\partial_s H\Big|_{\rho>0,s=0}=0.$

Then

$$\begin{aligned}2\pi\partial_{ss}H\Big|_{\rho=1,s=0}&=-\int\frac{A}{\sqrt{\beta+A^2}}\Big|_{s=0}\cdot\frac{xy^2}{\sqrt{\beta+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy\\&=-\int\frac{x^2y^2}{\beta+x^2}e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

<u>Calculation of $\partial_{\rho\rho}H$</u>. Clearly

$$2\pi\partial_\rho H=\int\sqrt{\beta+A^2}\frac{\beta\rho}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy.$$

Observe that

$$2\pi\partial_\rho H\Big|_{\rho=1,s=0}=\int\beta e^{-\frac{x^2+y^2}{2}}dxdy.$$

Then

$$\begin{aligned}2\pi\partial_{\rho\rho}H\Big|_{\rho=1,s=0}&=\int\sqrt{\beta+A^2}\Big|_{\rho=1,s=0}\frac{\beta}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy\\&\quad-\int\sqrt{\beta+A^2}\Big|_{\rho=1,s=0}\frac{\beta^2\rho^2}{(\beta\rho^2+x^2)^{\frac{3}{2}}}e^{-\frac{x^2+y^2}{2}}dxdy\\&=\int\frac{\beta x^2}{\beta+x^2}e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

<u>Calculation of $\partial_{\rho s}H$</u>. We have

$$\begin{aligned}2\pi\partial_{\rho s}H\Big|_{\rho=1,s=0}&=\int\frac{A}{\sqrt{\beta+A^2}}\Big|_{s=0}y\frac{\beta\rho}{\sqrt{\beta\rho^2+x^2}}e^{-\frac{x^2+y^2}{2}}dxdy\\&=\int\frac{\beta xy}{\beta+x^2}e^{-\frac{x^2+y^2}{2}}dxdy=0.\end{aligned}$$

Now we calculate the corresponding Hessian for $h=\frac{1}{2}(1+2\beta)\rho^2-\rho H$. Clearly

$$\begin{aligned}\partial_{\rho\rho}h\Big|_{\rho=1,s=0}&=1+2\beta-(\partial_{\rho\rho}H+2\partial_\rho H)\\&=1+2\beta-\frac{1}{2\pi}\int\Big(\frac{\beta x^2}{\beta+x^2}+2\beta\Big)e^{-\frac{x^2+y^2}{2}}dxdy\\&=\frac{1}{2\pi}\int\Big(1-\frac{\beta x^2}{\beta+x^2}\Big)e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

By Lemma A.8, this is clearly positive and has a lower bound depending only in terms of β.

On the other hand,

$$\partial_{\rho s}h\Big|_{\rho=1,s=0}=-\partial_s H-\partial_{\rho s}H=0.$$

Finally

$$\partial_{ss}h=-\partial_{ss}H=\frac{1}{2\pi}\int\frac{x^2y^2}{\beta+x^2}e^{-\frac{x^2+y^2}{2}}dxdy>0.$$

Thus, we complete the proof. $\square$

Lemma A.8. *For any $0<\beta<\infty$, we have*

$$\int\Big(1-\frac{\beta x^2}{\beta+x^2}\Big)e^{-\frac{x^2}{2}}dx>0.$$

Proof. For $0<\beta\le 1$, this is obvious. For $\beta>1$, denote $\epsilon=\frac{1}{\beta}$. Then

$$\begin{aligned}\tilde{h}(\epsilon)&=\int\Big(1-\frac{x^2}{1+\epsilon x^2}\Big)e^{-\frac{x^2}{2}}dx;\\\tilde{h}'(\epsilon)&=\int\frac{\epsilon x^4}{(1+\epsilon x^2)^2}e^{-\frac{x^2}{2}}dx>0.\end{aligned}$$

Clearly $\tilde{h}(0)=0$. Thus $\tilde{h}(\epsilon)>0$ for all $0<\epsilon<\infty$. $\square$

Lemma A.9. *Let $0<c_1<c_2<\infty$ be fixed. Consider for $\xi\in\mathbb{S}^{n-1}$, $u\in\mathbb{R}^n$ with $c_1\le\|u\|_2\le c_2$, the following:*

$$\begin{aligned}I_1=I_1(\xi,u)&=\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1))^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2,\\I_2=I_2(\xi,u)&=\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1))^{\frac{1}{2}}\cdot(a_k\cdot u)\cdot(a_k\cdot\xi).\end{aligned}$$

For any $0<\epsilon\le 1$, *if* $m\gtrsim n$, *then it holds with probability at least* $1-\mathcal{O}(m^{-2})$ *that*

$$\begin{aligned}&|I_1-\mathbb{E}I_1|\le\epsilon, \quad \forall\xi\in\mathbb{S}^{n-1}, \quad \forall c_1\le\|u\|_2\le c_2;\\ &|I_2-\mathbb{E}I_2|\le\epsilon, \quad \forall\xi\in\mathbb{S}^{n-1}, \quad \forall c_1\le\|u\|_2\le c_2.\end{aligned}$$

Proof. We first note that, in order to prove the statement for I_2, it suffices for us to prove the statement for I_1 under a more general condition (instead of $\xi\in\mathbb{S}^{n-1}$):

$$\|\xi\|_2\le c_3:=2+c_2.$$

The reason is as follows. By using the simple identity

$$(a_k\cdot(\xi+u))^2=(a_k\cdot\xi)^2+(a_k\cdot u)^2+2(a_k\cdot\xi)(a_k\cdot u),$$

we have

$$I_2(\xi,u)=\frac{1}{2}I_1(\xi+u,u)-\frac{1}{2}I_1(\xi,u)-I_3,$$

where

$$I_3=\frac{1}{m}\sum_{k=1}^{m}(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(a_k\cdot u)^2(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}.$$

Clearly I_3 is OK for union bounds and we have with high probability

$$|I_3-\mathbb{I}_3|\le\epsilon, \quad \forall c_1\le\|u\|_2\le c_2.$$

Thus to prove the statement for I_2 it suffices for us to prove it for I_1 uniformly in ξ with $\|\xi\|_2\le c_3$.

Next we observe that for $\xi\neq 0$ with $\|\xi\|_2\le c_3$, we have

$$\begin{aligned}|I_1(\xi,u)-\mathbb{E}I_1(\xi,u)|&\le\|\xi\|_2\Big|I_1\Big(\frac{\xi}{\|\xi\|_2},u\Big)-\mathbb{E}I_1\Big(\frac{\xi}{\|\xi\|_2},u\Big)\Big|\\ &\le c_3\Big|I_1\Big(\frac{\xi}{\|\xi\|_2},u\Big)-\mathbb{E}I_1\Big(\frac{\xi}{\|\xi\|_2},u\Big)\Big|.\end{aligned}$$

Thus it suffices for us to prove the statement for I_1 under the original assumption $\xi\in\mathbb{S}^{n-1}$.

Now let $\phi\in C_c^\infty(\mathbb{R})$ be such that $0\le\phi(x)\le 1$ for all x, $\phi(x)=1$ for $|x|\le 1$ and $\phi(x)=0$ for $|x|\ge 2$. Let $\delta>0$ be a sufficiently small constant. The needed smallness will be specified later. We write (below $\langle x\rangle=(1+x^2)^{\frac{1}{2}}$)

$$\begin{aligned}I_1=&\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\phi\Big(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle}\Big)\\&+\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\Big(1-\phi\Big(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle}\Big)\Big)\\=:&I_{1,a}+I_{1,b}.\end{aligned}$$

Estimate of $I_{1,a}$. Let $K=\delta^{-\frac{1}{9}}$. Then

$$\begin{aligned}|I_{1,a}|\le&\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\phi\Big(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle}\Big)\phi\Big(\frac{a_k\cdot\xi}{K}\Big)\\&+\frac{1}{m}\sum_{k=1}^m(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\phi\Big(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle}\Big)\cdot\Big(1-\phi\Big(\frac{a_k\cdot\xi}{K}\Big)\Big)\\\lesssim&K^2\frac{1}{m}\sum_{k=1}^m(1+(a_k\cdot e_1)^2)^{\frac{1}{2}}\phi\Big(\frac{a_k\cdot u}{\delta\langle 2K\rangle}\Big)+\frac{1}{m}\sum_{k=1}^m(1+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\Big(1-\phi\Big(\frac{a_k\cdot\xi}{K}\Big)\Big)\\\lesssim&K^5\frac{1}{m}\sum_{k=1}^m\phi\Big(\frac{a_k\cdot u}{\delta\langle 2K\rangle}\Big)+\frac{1}{m}\sum_{k=1}^m(1+(a_k\cdot e_1)^2)\cdot K^{-1}+K\cdot\frac{1}{m}\sum_{k=1}^m(a_k\cdot\xi)^4\cdot\Big(1-\phi\Big(\frac{a_k\cdot\xi}{K}\Big)\Big).\end{aligned}$$

Clearly for sufficiently small δ, we have

$$\mathbb{E}|I_{1,a}|\le\frac{\epsilon}{10}.$$

Furthermore, with probability at least $1-O(m^{-2})$, we have

$$|I_{1,a}|\le\frac{\epsilon}{10},\quad\forall c_1\le\|u\|_2\le c_2,\quad\forall\xi\in\mathbb{S}^{n-1}.$$

Estimate of $I_{1,b}$. Thanks to the cut-off $1-\phi(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle})$, we have $|a_k\cdot u|\gtrsim\langle a_k\cdot\xi\rangle$ on its support. It is then easy to check that the summands in $I_{1,b}$ are sub-exponential random variables. It remains for us to check the union bound.

To this end, take u, $\tilde u$ with $c_1\le\|u\|_2$, $\|\tilde u\|_2\le c_2$, and ξ, $\tilde\xi\in\mathbb{S}^{n-1}$. Then clearly

$$\begin{aligned}&\Bigg|(\beta|u|^2+(a_k\cdot u)^2)^{-\frac{3}{2}}(\beta|u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\xi)^2\cdot\Big(1-\phi\Big(\frac{a_k\cdot u}{\delta\langle a_k\cdot\xi\rangle}\Big)\Big)\\&\quad-(\beta|\tilde u|^2+(a_k\cdot\tilde u)^2)^{-\frac{3}{2}}(\beta|\tilde u|^2+(a_k\cdot e_1)^2)^{\frac{1}{2}}\cdot(a_k\cdot\tilde\xi)^2\cdot\Big(1-\phi\Big(\frac{a_k\cdot\tilde u}{\delta\langle a_k\cdot\tilde\xi\rangle}\Big)\Big)\Bigg|\\\lesssim&(1+|a_k\cdot e_1|)\cdot\Big(\|u-\tilde u\|_2+|a_k\cdot(u-\tilde u)|+|a_k\cdot(\xi-\tilde\xi)|\Big).\end{aligned}$$

Here in the above derivation we have used the fact that the function (it differs from the actual one by some minor change of parameters)

$$G(t,s)=\langle t\rangle^{-3}s^2\Big(1-\phi\Big(\frac{t}{\langle s\rangle}\Big)\Big)$$

satisfies

$$|G(t,s)-G(\tilde{t},\tilde{s})|\lesssim|t-\tilde{t}|+|s-\tilde{s}|.$$

It is then clear that $I_{1,b}$ is OK for union bounds and we have with high probability

$$|I_{1,b}-\mathbb{E}I_{1,b}|\le\frac{\epsilon}{10},\quad \forall c_1\le\|u\|_2\le c_2,\quad \forall\xi\in S^{n-1}.$$

The desired estimate for I_1 then easily follows. □

Lemma A.10. *Let $0<c_1<c_2<\infty$ be fixed. Consider*

$$f_0(u)=-\frac{1}{m}\sum_{k=1}^{m}\sqrt{\beta|u|^2+(a_k\cdot u)^2}\sqrt{\beta|u|^2+(a_k\cdot e_1)^2}.$$

For any $0<\epsilon\le1$, if $m\gtrsim n$, then it holds with probability at least $1-\mathcal{O}(m^{-2})$ that

$$\Big|\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f_0)(u)-\sum_{i,j=1}^{n}\xi_i\xi_j\mathbb{E}(\partial_{ij}f_0)(u)\Big|\le\epsilon,\quad \forall\xi\in\mathbb{S}^{n-1},\quad \forall c_1\le\|u\|_2\le c_2.$$

Proof. To simplify the notation, write a_k as a, and denote

$$\begin{aligned} &A=\beta|u|^2+(a\cdot u)^2, && B=\beta|u|^2+(a\cdot e_1)^2;\\ &\partial_iA=2\beta u_i+2(a\cdot u)a_i, && \partial_{ij}A=2\beta\delta_{ij}+2a_ia_j;\\ &\partial_iB=2\beta u_i, && \partial_{ij}B=2\beta\delta_{ij}. \end{aligned}$$

We need to compute $\partial_{ij}\tilde{F}$ for

$$\tilde{F}=A^{\frac12}B^{\frac12}.$$

Clearly

$$\begin{aligned} \partial_i\tilde{F}=&\frac12A^{-\frac12}\partial_iAB^{\frac12}+\frac12B^{-\frac12}\partial_iBA^{\frac12};\\ \partial_{ij}\tilde{F}=&-\frac14A^{-\frac32}\partial_iA\partial_jAB^{\frac12}+\frac12A^{-\frac12}\partial_{ij}AB^{\frac12}+\frac12A^{-\frac12}\partial_iA\frac12B^{-\frac12}\partial_jB\\ &-\frac14B^{-\frac32}\partial_jB\partial_iBA^{\frac12}+\frac12B^{-\frac12}\partial_{ij}BA^{\frac12}+\frac14B^{-\frac12}A^{-\frac12}\partial_iB\partial_jA. \end{aligned}$$

We then have

$$\begin{aligned}
&\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f_0)(u)\\
=&\frac{1}{4}\cdot\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}|\xi\cdot\nabla A_k|^2B_k^{\frac{1}{2}} &\text{(A.5)}\\
&+\frac{1}{4}\cdot\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{3}{2}}|\xi\cdot\nabla B_k|^2A_k^{\frac{1}{2}} &\text{(A.6)}\\
&-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{1}{2}}\langle\xi,(\nabla^2A_k)\xi\rangle B_k^{\frac{1}{2}} &\text{(A.7)}\\
&-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}(\nabla A_k\cdot\xi)(\nabla B_k\cdot\xi) &\text{(A.8)}\\
&-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{1}{2}}A_k^{\frac{1}{2}}\langle\xi,(\nabla^2B_k)\xi\rangle, &\text{(A.9)}
\end{aligned}$$

where

$$A_k=\beta|u|^2+(a_k\cdot u)^2,\quad B_k=\beta|u|^2+(a_k\cdot e_1)^2,$$

and we have denoted

$$\langle\xi,(\nabla^2A_k)\xi\rangle=\sum_{i,j=1}^{n}\xi_i\xi_j\partial_{ij}A_k.$$

Estimate of (A.9). We have

$$\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{1}{2}}A_k^{\frac{1}{2}}\langle\xi,(\nabla^2B_k)\xi\rangle=2\beta|\xi|^2\Big(\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{1}{2}}A_k^{\frac{1}{2}}\Big).$$

The summand consists of sub-exponential random variables and are clearly OK for union bounds. Thus with high probability, it holds that

$$\Big|\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{1}{2}}A_k^{\frac{1}{2}}-\text{mean}\Big|\le\frac{\epsilon}{100(1+2\beta)},\quad\forall c_1\le\|u\|_2\le c_2.$$

Thus the contribution of (A.9) is OK for us.
Estimate of (A.6). We have

$$\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{3}{2}}A_k^{\frac{1}{2}}|\xi\cdot\nabla B_k|^2=4\beta^2(\xi\cdot u)^2\Big(\frac{1}{m}\sum_{k=1}^{m}B_k^{-\frac{3}{2}}A_k^{\frac{1}{2}}\Big).$$

Again the summand consists of sub-exponential random variables and are clearly OK for union bounds. Thus with high probability, it holds that

$$\Big|\frac{1}{m}\sum_{k=1}^m B_k^{-\frac{3}{2}}A_k^{\frac{1}{2}}-\text{mean}\Big|\le\frac{\epsilon}{100(1+4\beta^2c_2^2)},\quad \forall c_1\le\|u\|_2\le c_2.$$

Thus the contribution of (A.6) is OK for us.

Estimate of (A.8). We have

$$\begin{aligned}
&\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}(\nabla A_k\cdot\xi)(\nabla B_k\cdot\xi)\\
=&\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}\cdot\Big(2\beta(u\cdot\xi)+2(a_k\cdot u)(a_k\cdot\xi)\Big)2\beta(\xi\cdot u)\\
=&4\beta^2(\xi\cdot u)^2\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}\\
&\quad+4\beta(\xi\cdot u)\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}(a_k\cdot\xi)(a_k\cdot u).
\end{aligned}\tag{A.10}$$

The first term is clearly under control and therefore we focus only on (A.10). For this observe that for any u, $\tilde{u}$ with $c_1\le\|u\|_2$, $\|\tilde{u}\|_2\le c_2$, ξ, $\tilde{\xi}\in\mathbb{S}^{n-1}$, it holds that

$$\begin{aligned}
&\Big|\frac{a_k\cdot u}{\sqrt{\beta|u|^2+|a_k\cdot u|^2}}-\frac{a_k\cdot\tilde{u}}{\sqrt{\beta|u|^2+|a_k\cdot\tilde{u}|^2}}\Big|\lesssim|a_k\cdot(u-\tilde{u})|,\\
&\Big|\frac{a_k\cdot\tilde{u}}{\sqrt{\beta|u|^2+|a_k\cdot\tilde{u}|^2}}-\frac{a_k\cdot\tilde{u}}{\sqrt{\beta|\tilde{u}|^2+|a_k\cdot\tilde{u}|^2}}\Big|\lesssim\|u-\tilde{u}\|_2,\\
&\Big|\frac{a_k\cdot u}{\sqrt{\beta|u|^2+|a_k\cdot u|^2}}\cdot\frac{a_k\cdot\xi}{\sqrt{\beta|u|^2+|a_k\cdot e_1|^2}}-\frac{a_k\cdot\tilde{u}}{\sqrt{\beta|\tilde{u}|^2+|a_k\cdot\tilde{u}|^2}}\cdot\frac{a_k\cdot\tilde{\xi}}{\sqrt{\beta|\tilde{u}|^2+|a_k\cdot e_1|^2}}\Big|\\
&\lesssim(|a_k\cdot(u-\tilde{u})|+\|u-\tilde{u}\|_2)|a_k\cdot\xi|+|a_k\cdot(\xi-\tilde{\xi})|.
\end{aligned}$$

Thus (A.10) is OK for union bounds and we have with high probability,

$$\begin{aligned}
&\Big|\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}(a_k\cdot\xi)(a_k\cdot u)-\text{mean}\Big|\\
\le&\frac{\epsilon}{200(1+4\beta c_2)},\quad\forall c_1\le\|u\|_2\le c_2,\quad\forall\xi\in\mathbb{S}^{n-1}.
\end{aligned}$$

Thus (A.8) is under control.

Estimate of (A.7). We have

$$
\begin{aligned}
&-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{1}{2}}\langle\xi,(\nabla^2 A_k)\xi\rangle B_k^{\frac{1}{2}}\\
=&-\beta|\xi|^2\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{1}{2}}B_k^{\frac{1}{2}}\\
&-\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{1}{2}}B_k^{\frac{1}{2}}(a_k\cdot\xi)^2. \qquad\qquad (A.11)
\end{aligned}
$$

The first term is clearly under control. Therefore we only need to treat (A.11). We shall treat it together with (A.12) below.

Estimate of (A.5). We have

$$
\begin{aligned}
&\frac{1}{4}\cdot\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}|\xi\cdot\nabla A_k|^2 B_k^{\frac{1}{2}}\\
=&\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}B_k^{\frac{1}{2}}(a_k\cdot u)^2(a_k\cdot\xi)^2 \qquad\qquad (A.12)\\
&+2\beta(\xi\cdot u)\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}B_k^{\frac{1}{2}}(a_k\cdot u)(a_k\cdot\xi) \qquad\qquad (A.13)\\
&+\beta^2(\xi\cdot u)^2\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}B_k^{\frac{1}{2}}. \qquad\qquad (A.14)
\end{aligned}
$$

Clearly (A.14) is perfectly under control. Now observe

$$
(A.11)+(A.12)=-\beta|u|^2\frac{1}{m}\sum_{k=1}^{m}A_k^{-\frac{3}{2}}B_k^{\frac{1}{2}}(a_k\cdot\xi)^2.
$$

One can then apply Lemma A.9 to get the desired estimate for this term as well as (A.13). □

Appendix B: technical estimates for Section 3

Lemma B.1. *Denote $X_j=a_j\cdot e_1$ and $Z_j=a_j\cdot\hat{u}$, where $\hat{u}\in\mathbb{S}^{n-1}$. For any $\epsilon>0$, there exists $R=R(\epsilon,\beta)>0$, such that if $m\gtrsim n$, then the following hold with high probability:*

$$
\frac{1}{m}\sum_{j=1}^{m}(\beta+Z_j^2)\sqrt{\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}}\le\epsilon,\quad \forall 0<\rho\le R,\quad \forall\hat{u}\in\mathbb{S}^{n-1}.
$$

Proof. We shall only sketch the proof. Choose $\phi\in C_c^\infty(\mathbb{R})$ such that $0\le\phi(x)\le 1$ for all x, $\phi(x)=1$ for $|x|\le 1$ and $\phi(x)=0$ for $|x|\ge 2$. Then

$$\begin{aligned}&(\beta+Z_j^2)\sqrt{\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}}\\ \le&(\beta+Z_j^2)\sqrt{\frac{\rho^2(\beta+2Z_j^2)}{\rho^2(\beta+Z_j^2)+X_j^2}}\phi\Big(\frac{Z_j}{K}\Big)+(\beta+Z_j^2)\cdot\sqrt{2}\cdot\Big(1-\phi\Big(\frac{Z_j}{K}\Big)\Big),\end{aligned}\tag{B.1}$$

where $K>0$ is a constant to be specified momentarily. Clearly by taking K sufficiently large, we have with high probability that

$$\frac{1}{m}\sum_{j=1}^m(\beta+Z_j^2)\cdot\sqrt{2}\cdot\Big(1-\phi\Big(\frac{Z_j}{K}\Big)\Big)\le\frac{\epsilon}{10},\quad\forall\hat{u}\in\mathbb{S}^{n-1}.$$

It then remains for us to deal with (B.1). Thanks to the smooth cut-off, we have

$$\begin{aligned}\text{(B.1)}\le&\rho C_{K,\beta}\cdot\frac{1}{\sqrt{\rho^2\beta+X_j^2}}\\ \le&\rho C_{K,\beta}\cdot\frac{1}{\eta}+E_{K,\beta}\cdot\phi\Big(\frac{X_j}{\eta}\Big),\end{aligned}$$

where $C_{K,\beta}>0$, $E_{K,\beta}>0$ are constants depending only on K and β. We first choose $\eta>0$ sufficiently small such that with high probability,

$$E_{K,\beta}\Big|\frac{1}{m}\sum_{j=1}^m\phi\Big(\frac{X_j}{\eta}\Big)\Big|\le\frac{\epsilon}{10}.$$

Then the desired result follows by taking ρ sufficiently small. $\square$

Lemma B.2. *Let $\gamma_1>0$, $\gamma_2>0$ and $\gamma_3\ge 0$. Consider*

$$g(\theta_0)=\int_0^\pi\sqrt{\gamma_1+\gamma_2\cos^2(\theta-\theta_0)+\gamma_3\sin^2\theta}\sqrt{\gamma_1+2\gamma_3\sin^2\theta}d\theta.$$

Then

$$g'(\theta_0)\ge 0,\quad\forall\theta_0\in\Big[0,\frac{\pi}{2}\Big).$$

Furthermore, if $\gamma_1\sim 1$, $\gamma_2\sim 1$, $\gamma_3\ge 0$, then

$$g'(\theta_0)\gtrsim\frac{1}{1+\gamma_3}\sin 2\theta_0.$$

In particular we have

$$g''(0)\gtrsim\frac{1}{1+\gamma_3}.$$

Remark B.1. There exists a subtle balance of coefficients in the expression of $g(\theta_0)$ without which we cannot have the positivity of g'. As a counter-example, consider

$$f(s,b)=\int_0^{\pi}(1+b\cos^2(\theta-s)+2\sin^2\theta)^{\frac{1}{2}}(1+\sin^2\theta)^{\frac{1}{2}}d\theta.$$

One can check that $\partial_s f(s,b)<0$ for $b<1.99$ and $\partial_s f(s,b)>0$ for some $b\geq 2$ and s.

Proof of Lemma B.2. Clearly

$$\begin{aligned}
g'(\theta_0)&=\gamma_2\int_0^{\pi}\sqrt{\frac{\gamma_1+2\gamma_3\sin^2\theta}{\gamma_1+\gamma_2\cos^2(\theta-\theta_0)+\gamma_3\sin^2\theta}}\sin2(\theta-\theta_0)d\theta\\
&=\gamma_2\int_0^{\pi}\sqrt{\frac{\gamma_1+2\gamma_3\sin^2(\theta+\theta_0)}{\gamma_1+\gamma_2\cos^2\theta+\gamma_3\sin^2(\theta+\theta_0)}}\sin2\theta d\theta\\
&=\gamma_2\int_0^{\frac{\pi}{2}}\left(\sqrt{\frac{\gamma_1+2\gamma_3\sin^2(\theta+\theta_0)}{\gamma_1+\gamma_2\cos^2\theta+\gamma_3\sin^2(\theta+\theta_0)}}-\sqrt{\frac{\gamma_1+2\gamma_3\sin^2(\theta-\theta_0)}{\gamma_1+\gamma_2\cos^2\theta+\gamma_3\sin^2(\theta-\theta_0)}}\right)\sin2\theta d\theta.
\end{aligned}$$

Clearly for $\theta,\theta_0\in[0,\frac{\pi}{2})$, we have

$$\sin(\theta+\theta_0)\geq|\sin(\theta-\theta_0)|.$$

The non-negativity of g' then follows from the monotonicity of the function (below $a\geq 1$ is a constant)

$$\tilde{g}(z)=\frac{\gamma_1+2\gamma_3 z}{a\gamma_1+\gamma_3 z}=2-\frac{(2a-1)\gamma_1}{\gamma_3 z+a\gamma_1},\quad z\geq 0.$$

Next if $\gamma_1,\gamma_2\sim 1,\gamma_3\geq 0$, then clearly (note that $a=1+\frac{\gamma_2}{\gamma_1}\cos^2\theta\geq 1$, $a\sim 1$)

$$\tilde{g}'(z)\gtrsim\frac{1}{1+\gamma_3},\quad \forall z\in[0,1].$$

Thus

$$\begin{aligned}
g'(\theta_0)&\gtrsim\frac{1}{1+\gamma_3}\int_0^{\frac{\pi}{2}}(\sin^2(\theta+\theta_0)-\sin^2(\theta-\theta_0))\sin2\theta d\theta\\
&\gtrsim\frac{1}{1+\gamma_3}\int_0^{\frac{\pi}{2}}\sin^2(2\theta)d\theta\sin2\theta_0\\
&\gtrsim\frac{1}{1+\gamma_3}\sin2\theta_0.
\end{aligned}$$

Since

$$g''(0)=\lim_{\theta_0\to 0+}\frac{g'(\theta_0)}{\theta_0},$$

the estimate for $g''(0)$ easily follows. □

Proof of Lemma 3.3. Clearly

$$\begin{aligned}h_\infty(\rho,t)=&\mathbb{E}\sqrt{\beta\rho^2+\rho^2X_t^2+X_1^2}\sqrt{\beta\rho^2+2X_1^2}\\=&\frac{1}{2\pi}\int_{\mathbb{R}^2}\sqrt{\beta\rho^2+\rho^2(tx+\sqrt{1-t^2}y)^2+x^2}\sqrt{\beta\rho^2+2x^2}e^{-\frac{x^2+y^2}{2}}dxdy.\end{aligned}$$

Since $\rho\sim 1$, it is easy to check that

$$\sup_{|t|\le 1-\eta_0}(|\partial_t h_\infty(\rho,t)|+|\partial_{tt}h_\infty(\rho,t)|+|\partial_{ttt}h_\infty(\rho,t)|\lesssim 1.$$

To show the lower bound on $|\partial_t h_\infty(\rho,t)|$, observe that $h_\infty(\rho,t)$ is an even function of t. Thus without loss of generality we assume $0\le t<1$. Now let $t=\sin\theta_0$ with $\theta_0\in[0,\frac{\pi}{2})$. By using polar coordinates, we obtain

$$\begin{aligned}h_\infty(\rho,t)=&\frac{1}{2\pi}\int_0^\infty\int_0^{2\pi}\sqrt{\beta\rho^2+\rho^2\cos^2(\theta-\theta_0)+r^2\sin^2\theta}\sqrt{\beta\rho^2+2r^2\sin^2\theta}e^{-\frac{r^2}{2}}rd\theta dr\\=&\frac{1}{\pi}\int_0^\infty\int_0^{\pi}\sqrt{\beta\rho^2+\rho^2\cos^2(\theta-\theta_0)+r^2\sin^2\theta}\sqrt{\beta\rho^2+2r^2\sin^2\theta}e^{-\frac{r^2}{2}}rd\theta dr.\end{aligned}$$

Observe that

$$\partial_{\theta_0}\Big(h_\infty(\rho,\sin\theta_0)\Big)=(\partial_t h_\infty)(\rho,t)\Big|_{t=\sin\theta_0}\cos\theta_0. \tag{B.2}$$

By Lemma B.2 (note that $\gamma_3=r^2$) and integrating in r, we then obtain

$$\partial_t h_\infty(\rho,t)\gtrsim t,\quad \forall 0\le t<1.$$

Finally to show that $\partial_{tt}h_\infty(\rho,t)\gtrsim 1$ for $|t|\ll 1$, it suffices for us to show (since $|\partial_{ttt}h_\infty(\rho,t)|\lesssim 1$ for $|t|\ll 1$)

$$\partial_{tt}h_\infty(\rho,0)\gtrsim 1.$$

By using (B.2), we only need to check

$$\partial_{\theta_0\theta_0}\Big(h_\infty(\rho,\sin\theta_0)\Big)\Big|_{\theta_0=0}\gtrsim 1.$$

This again follows from Lemma B.2. □

Lemma B.3. *Suppose $\phi_1:\mathbb{R}\to\mathbb{R}$, $\phi_2:\mathbb{R}\to\mathbb{R}$ are C^1 functions such that*

$$\max_{|z|\le L}(|\phi_1(z)|+|\phi_1'(z)|+|\phi_2(z)|+|\phi_2'(z)|)\le C_{L,\phi_1,\phi_2},$$

where $C_{L,\phi_1,\phi_2}>0$ is finite for each finite L.

Suppose $0<c_1<c_2<\infty$ and $\phi_3:(\frac{c_1}{2},2c_2)\to\mathbb{R}$ is a smooth function such that

$$\sup_{\frac{c_1}{2}<|z|<2c_2}(|\phi_3(z)|+|\phi_3'(z)|)\le C_{c_1,c_2,\phi_3},$$

where $C_{c_1,c_2,\phi_3}>0$ depends only on c_1, c_2 and ϕ_3.

Let $(d_{ij})_{1\le i\le 2,1\le j\le 3}$ be given constants and consider

$$\begin{aligned}&I(u,w,v)\\=&\frac{1}{m}\sum_{j=1}^m\phi_1\Big(\frac{d_{11}|u|^2+d_{12}(a_j\cdot e_1)^2+d_{13}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\phi_2\Big(\frac{d_{21}|u|^2+d_{22}(a_j\cdot e_1)^2+d_{23}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\\&\cdot\phi_3(\|u\|_2)(a_j\cdot w)(a_j\cdot v),\quad u\in\mathbb{R}^n,\quad w,v\in\mathbb{S}^{n-1}.\end{aligned}$$

Then for any $0<\epsilon\le 1$, if $m\gtrsim n$, then the following hold with high probability:

$$|I(u,w,v)-\mathbb{E}I(u,w,v)|\le\epsilon,\quad\forall w,v\in\mathbb{S}^{n-1},\quad\forall c_1\le\|u\|_2\le c_2.$$

Proof. We first note that, by using a polarization argument and scaling (cf. the beginning part of the proof of Lemma A.9), it suffices for us to prove the statement for $I(u,w,w)$ uniformly in $w\in\mathbb{S}^{n-1}$ and $u\in\mathbb{R}^n$ with $c_1\le\|u\|_2\le c_2$.

Now let $\phi\in C_c^\infty(\mathbb{R})$ such that $0\le\phi(x)\le 1$ for all x, $\phi(x)=1$ for $|x|\le 1$ and $\phi(x)=0$ for $|x|\ge 2$. Let $\delta>0$ be a sufficiently small constant. The smallness of δ will be specified momentarily. Then

$$\begin{aligned}&|I_1(u,w)|\\=&\Big|\frac{1}{m}\sum_{j=1}^m\phi_1\Big(\frac{d_{11}|u|^2+d_{12}(a_j\cdot e_1)^2+d_{13}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\phi_2\Big(\frac{d_{21}|u|^2+d_{22}(a_j\cdot e_1)^2+d_{23}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\\&\cdot\phi_3(\|u\|_2)(a_j\cdot w)^2\phi\Big(\frac{a_j\cdot u}{\delta\langle a_j\cdot w\rangle}\Big)\Big|\\\lesssim&\frac{1}{m}\sum_{j=1}^m(a_j\cdot w)^2\phi\Big(\frac{a_j\cdot u}{\delta\langle a_j\cdot w\rangle}\Big)\\\lesssim&\frac{1}{m}\sum_{j=1}^m(a_j\cdot w)^2\Big(1-\phi(2\delta^{\frac18}(a_j\cdot w))\Big)+\frac{1}{m}\sum_{j=1}^m\delta^{-\frac14}\phi\Big(\frac{a_j\cdot u}{\delta\langle\delta^{-\frac18}\rangle}\Big).\end{aligned}$$

The expectation of the above two terms are clearly small if we take $\delta>0$ sufficiently small. Moreover they are clearly OK for union bounds and can be made small in high probability. Thus for sufficiently small δ, if $m\gtrsim n$, then with high probability we have

$$|I_1(u,w)-\mathbb{E}I_1(u,w)|\le\frac{\epsilon}{3},\quad \forall w\in\mathbb{S}^{n-1},\quad \forall c_1\le\|u\|_2\le c_2.$$

We now fix δ and deal with the main term

$$\begin{aligned}&I_2(u,w)\\ =&\frac{1}{m}\sum_{j=1}^m\phi_1\Big(\frac{d_{11}|u|^2+d_{12}(a_j\cdot e_1)^2+d_{13}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\phi_2\Big(\frac{d_{21}|u|^2+d_{22}(a_j\cdot e_1)^2+d_{23}(a_j\cdot u)^2}{\beta|u|^2+(a_j\cdot u)^2+(a_j\cdot e_1)^2}\Big)\\ &\cdot\phi_3(\|u\|_2)(a_j\cdot w)^2\cdot\left(1-\phi\Big(\frac{a_j\cdot u}{\delta\langle a_j\cdot w\rangle}\Big)\right)\\ =&\frac{1}{m}\sum_{j=1}^m H(\|u\|_2,a_j\cdot u,a_j\cdot w,a_j\cdot e_1),\end{aligned}$$

where

$$\begin{aligned}&H(s,z,y,b)\\ =&\phi_1\Big(\frac{d_{11}s^2+d_{12}b^2+d_{13}z^2}{\beta s^2+z^2+b^2}\Big)\phi_2\Big(\frac{d_{21}s^2+d_{22}b^2+d_{23}z^2}{\beta s^2+z^2+b^2}\Big)\cdot\phi_3(s)y^2\left(1-\phi\Big(\frac{z}{\delta\langle y\rangle}\Big)\right).\end{aligned}$$

The main point is to check the union bounds. Note that $s=\|u\|_2\sim 1$. We have

$$\begin{aligned}|\partial_s H(s,z,y,b)|&\lesssim y^2;\\ |\partial_z H(s,z,y,b)|&\lesssim |y|;\\ |\partial_y H(s,z,y,b)|&\lesssim |y|.\end{aligned}$$

Thus for $c_1\le\|u\|_2,\|\tilde u\|_2\le c_2$, $w,\tilde w\in\mathbb{S}^{n-1}$, we have

$$\begin{aligned}&\Big|H(\|u\|_2,a_j\cdot u,a_j\cdot w,a_j\cdot e_1)-H(\|\tilde u\|_2,a_j\cdot\tilde u,a_j\cdot\tilde w,a_j\cdot e_1)\Big|\\ \lesssim&\|u-\tilde u\|_2(|a_j\cdot w|^2)+|a_j\cdot(u-\tilde u)||a_j\cdot w|+|a_j\cdot(w-\tilde w)|(|a_j\cdot w|+|a_j\cdot\tilde w|).\end{aligned}$$

Clearly then the union bounds hold for I_2. Thus for $m\gtrsim n$, with high probability it holds that

$$|I_2(u,w)-\mathbb{E}I_2(u,w)|\le\frac{\epsilon}{3},\quad \forall w\in\mathbb{S}^{n-1},\quad \forall c_1\le\|u\|_2\le c_2.$$

The desired estimate for $I(u,w,w)$ then easily follows. $\square$

Consider

$$h(\rho,t,e^{\perp})=\frac{1}{m}\sum_{j=1}^{m}\sqrt{\beta\rho^2+\rho^2(a_j\cdot\hat{u})^2+X_j^2}\cdot\sqrt{\beta\rho^2+2X_j^2},$$

where

$$\begin{aligned}&X_j=a_j\cdot e_1, \quad u=\rho\hat{u}, \quad 0<c_1\le\rho\le c_2<\infty;\\&\hat{u}=te_1+\sqrt{1-t^2}e^{\perp}, \quad |t|<1, \quad e^{\perp}\cdot e_1=0, \quad e^{\perp}\in\mathbb{S}^{n-1}.\end{aligned}$$

Here we take $c_1>0$, $c_2>0$ as two fixed constants. The main point is that $\rho\sim 1$. We consider h in the regime

$$|t|\le 1-\epsilon_0,$$

where $0<\epsilon_0\ll 1$ is fixed.

Lemma B.4. *Let $0<\epsilon_0\ll 1$ be fixed. For any $0<\epsilon\le 1$, if $m\gtrsim n$, then with high probability it holds that*

$$|\partial_t h-\mathbb{E}\partial_t h|+|\partial_{tt}h-\mathbb{E}\partial_{tt}h|\le\epsilon, \quad \forall|t|\le 1-\epsilon_0, \quad e^{\perp}\cdot e_1=0, \quad e^{\perp}\in\mathbb{S}^{n-1}, \quad c_1\le\rho\le c_2.$$

Proof of Lemma A.4. Denote $Y_j=a_j\cdot e^{\perp}$ and

$$Z_j=a_j\cdot\hat{u}=tX_j+\sqrt{1-t^2}Y_j.$$

Clearly

$$\begin{aligned}&\frac{d}{dt}Z_j=X_j-\frac{t}{\sqrt{1-t^2}}Y_j;\\&\frac{d^2}{dt^2}Z_j=-(1-t^2)^{-\frac{3}{2}}Y_j.\end{aligned}$$

Using

$$Y_j=(1-t^2)^{-\frac{1}{2}}(Z_j-tX_j),$$

we obtain

$$\begin{aligned}&\frac{d}{dt}Z_j=\frac{1}{1-t^2}X_j-\frac{t}{1-t^2}Z_j;\\&\frac{d^2}{dt^2}Z_j=(1-t^2)^{-2}(tX_j-Z_j).\end{aligned}$$

Therefore

$$\begin{aligned}
\partial_t h = & \frac{1}{1-t^2}\cdot\frac{1}{m}\sum_{j=1}^m \sqrt{\frac{\beta|u|^2+2X_j^2}{\beta|u|^2+(a_j\cdot u)^2+X_j^2}}\|u\|_2\cdot(a_j\cdot u)X_j \\
& -\frac{t}{1-t^2}\cdot\frac{1}{m}\sum_{j=1}^m \sqrt{\frac{\beta|u|^2+2X_j^2}{\beta|u|^2+(a_j\cdot u)^2+X_j^2}}(a_j\cdot u)^2 \\
=: & \frac{1}{1-t^2}H_1-\frac{t}{1-t^2}H_2.
\end{aligned}$$

By Lemma B.3, it holds with high probability that

$$|H_1-\mathbb{E}H_1|+|H_2-\mathbb{E}H_2|\le(1-\epsilon_0^2)\cdot\frac{\epsilon}{3}, \quad \forall\hat{u}\in\mathbb{S}^{n-1}, \quad c_1\le\|u\|_2\le c_2, \quad |t|\le 1-\epsilon_0.$$

The desired estimate for $\partial_t h$ then easily follows.

To compute $\partial_{tt}h$, we shall denote

$$\begin{aligned}
A_j &= \beta\rho^2+\rho^2 Z_j^2+X_j^2=\beta|u|^2+(a_j\cdot u)^2+X_j^2; \\
B_j &= \beta\rho^2+2X_j^2=\beta|u|^2+2X_j^2.
\end{aligned}$$

Then

$$\begin{aligned}
\partial_{tt}h = & -\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{3}{2}}B_j^{\frac{1}{2}}\Big(\rho^2 Z_j\frac{d}{dt}Z_j\Big)^2+\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{1}{2}}\sqrt{B_j}\rho^2\cdot\Big(\Big(\frac{d}{dt}Z_j\Big)^2+Z_j\frac{d^2}{dt^2}Z_j\Big) \\
= & -(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{3}{2}}B_j^{\frac{1}{2}}\|u\|_2^2(a_j\cdot u)^2X_j^2+2t(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{3}{2}}B_j^{\frac{1}{2}}\|u\|_2(a_j\cdot u)^3X_j \\
& -t^2(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{3}{2}}B_j^{\frac{1}{2}}(a_j\cdot u)^4+(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{1}{2}}B_j^{\frac{1}{2}}\|u\|_2^2X_j^2 \\
& -t(1-t^2)^{-2}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{1}{2}}B_j^{\frac{1}{2}}\|u\|_2X_j(a_j\cdot u)-(1-t^2)^{-1}\frac{1}{m}\sum_{j=1}^m A_j^{-\frac{1}{2}}B_j^{\frac{1}{2}}(a_j\cdot u)^2.
\end{aligned}$$

It is then a bit tedious but not difficult to verify that the above terms can be treated with the help of Lemma B.3. Thus with high probability it holds that

$$|\partial_{tt}h-\mathbb{E}\partial_{tt}h|\le\frac{\epsilon}{5}, \quad \forall\hat{u}\in\mathbb{S}^{n-1}, \quad c_1\le\|u\|_2\le c_2, \quad |t|\le 1-\epsilon_0.$$

Thus, we complete the proof. $\square$

Lemma B.5. *Let $X\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$ be independent. Define*

$$\begin{aligned}H(\rho,s)&=\mathbb{E}\sqrt{\beta\rho^2+\rho^2(\sqrt{1-s^2}X+sY)^2+X^2}\sqrt{\beta\rho^2+2X^2};\\h(\rho,s)&=\frac{1}{2}(1+2\beta)\rho^2-H(\rho,s).\end{aligned}$$

Then it holds that

$$\sup_{|\rho-1|\ll1,|s|\ll1}\sum_{j=1}^{3}(|\partial^j H|+|\partial^j h|)\lesssim 1,$$

where $\partial=\partial_\rho$ or ∂_s.

Proof. For $H(\rho,s)$, this is obvious since the integrand inside the expectation is smooth. The estimate for $h(\rho,s)$ also follows easily. $\square$

Lemma B.6 (Calculation of $\partial^2 h$ at $(\rho=1,\ s=0)$)**.** *Let*

$$\begin{aligned}H(\rho,s)&=\mathbb{E}\sqrt{\beta\rho^2+\rho^2(\sqrt{1-s^2}X+sY)^2+X^2}\sqrt{\beta\rho^2+2X^2};\\h(\rho,s)&=\frac{1}{2}(1+2\beta)\rho^2-H(\rho,s).\end{aligned}$$

Then at $\rho=1$, $s=0$, we have

$$\begin{aligned}&(\partial_{\rho\rho}H)(1,0)=\gamma_1>0,\quad (\partial_{\rho s}H)(\rho,0)=0,\quad \forall\rho>0;\\&(\partial_{ss}H)(1,0)=-\gamma_2<0;\\&(\partial_s h)(\rho,0)=0,\quad \forall\rho>0,\quad (\partial_\rho h)(1,0)=0;\\&(\partial_{\rho\rho}h)(1,0)=\gamma_3>0,\quad (\partial_{\rho s}h)(\rho,0)=0,\quad \forall\rho>0;\\&(\partial_{ss}h)(1,0)=\gamma_4>0,\end{aligned}$$

where $\gamma_i>0$, $i=1,\cdots,4$ are constants depending on β.

Proof. Firstly by using parity it is easy to check that $(\partial_s H)(\rho,0)=0$ for any $\rho>0$. It follows easily that $(\partial_{\rho s}h)(\rho,0)=(\partial_{\rho s}H)(\rho,0)=0$ for any $\rho>0$. It is also easy to check that

$$\begin{aligned}(\partial_\rho H)(1,0)&=\partial_\rho\mathbb{E}\Big(\sqrt{\beta\rho^2+(\rho^2+1)X^2}\sqrt{\beta\rho^2+2X^2}\Big)\Big|_{\rho=1}\\&=\mathbb{E}(2\beta+X^2)=2\beta+1.\end{aligned}$$

Clearly $(\partial_\rho h)(1,0)=0$. One should note that we can also deduce this directly (and easily) from the fact that the original loss function attains a minimum at $u=e_1$.

Calculation of $\partial_{ss}H$. By a tedious computation, we have

$$\begin{aligned}2\pi(\partial_{ss}H)(1,0)&=2\pi\partial_{ss}\Big(\mathbb{E}\sqrt{\beta+X^2+(\sqrt{1-s^2}X+sY)^2}\sqrt{\beta+2X^2}\Big)\Big|_{s=0}\\&=\int_{\mathbb{R}^2}\frac{-2x^4-\beta x^2+(\beta+x^2)y^2}{\beta+2x^2}e^{-\frac{x^2+y^2}{2}}dxdy\\&=\int_{\mathbb{R}}\frac{-2x^4-\beta x^2+\beta+x^2}{\beta+2x^2}e^{-\frac{x^2}{2}}dx\\&=\int_{\mathbb{R}}\Big(-x^2\frac{\beta+1+2x^2}{\beta+2x^2}+1\Big)e^{-\frac{x^2}{2}}dx\\&=-\int_{\mathbb{R}}\frac{x^2}{\beta+2x^2}e^{-\frac{x^2}{2}}dx<0.\end{aligned}$$

Calculation of $\partial_{\rho\rho}H$. By a tedious computation, we have

$$\begin{aligned}(\partial_{\rho\rho}H)(1,0)&=\partial_{\rho\rho}\Big(\mathbb{E}\sqrt{\beta\rho^2+(\rho^2+1)X^2}\sqrt{\beta\rho^2+2X^2}\Big)\Big|_{\rho=1}\\&=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{2\beta^2+5\beta x^2+x^4}{\beta+2x^2}e^{-\frac{x^2}{2}}dx\\&=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(2\beta+\frac{\beta+x^2}{\beta+2x^2}x^2)e^{-\frac{x^2}{2}}dx.\end{aligned}$$

It follows that

$$\begin{aligned}(\partial_{\rho\rho}h)(1,0)&=1+2\beta-(\partial_{\rho\rho}H)(1,0)\\&=-\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{x^4}{\beta+2x^2}e^{-\frac{x^2}{2}}dx<0.\end{aligned}$$

This completes the proof. □

Lemma B.7. *Let $0<c_1<c_2<\infty$ be fixed. Consider*

$$f_0(u)=-\frac{1}{m}\sum_{k=1}^{m}\sqrt{\beta|u|^2+(a_k\cdot u)^2+(a_k\cdot e_1)^2}\sqrt{\beta|u|^2+2(a_k\cdot e_1)^2}.$$

For any $0<\epsilon\leq 1$, if $m\gtrsim n$, then it holds with high probability that

$$\Big|\sum_{i,j=1}^{n}\xi_i\xi_j(\partial_{ij}f_0)(u)-\sum_{i,j=1}^{n}\xi_i\xi_j\mathbb{E}(\partial_{ij}f_0)(u)\Big|\leq\epsilon,\quad\forall\xi\in\mathbb{S}^{n-1},\quad\forall c_1\leq\|u\|_2\leq c_2.$$

Proof. To simplify the notation, write a_k as a, and denote

$$\begin{aligned} &A=\beta|u|^2+(a\cdot u)^2+(a\cdot e_1)^2, && B=\beta|u|^2+2(a\cdot e_1)^2;\\ &\partial_i A=2\beta u_i+2(a\cdot u)a_i, && \partial_{ij}A=2\beta\delta_{ij}+2a_ia_j;\\ &\partial_i B=2\beta u_i, && \partial_{ij}B=2\beta\delta_{ij}. \end{aligned}$$

We need to compute $\partial_{ij}\tilde{F}$ for

$$\tilde{F}=A^{\frac{1}{2}}B^{\frac{1}{2}}.$$

Clearly

$$\begin{aligned} \partial_i\tilde{F}=&\frac{1}{2}A^{-\frac{1}{2}}\partial_iAB^{\frac{1}{2}}+\frac{1}{2}B^{-\frac{1}{2}}\partial_iBA^{\frac{1}{2}};\\ \partial_{ij}\tilde{F}=&-\frac{1}{4}A^{-\frac{3}{2}}\partial_iA\partial_jAB^{\frac{1}{2}}+\frac{1}{2}A^{-\frac{1}{2}}\partial_{ij}AB^{\frac{1}{2}}+\frac{1}{2}A^{-\frac{1}{2}}\partial_iA\frac{1}{2}B^{-\frac{1}{2}}\partial_jB\\ &-\frac{1}{4}B^{-\frac{3}{2}}\partial_jB\partial_iBA^{\frac{1}{2}}+\frac{1}{2}B^{-\frac{1}{2}}\partial_{ij}BA^{\frac{1}{2}}+\frac{1}{4}B^{-\frac{1}{2}}A^{-\frac{1}{2}}\partial_iB\partial_jA. \end{aligned}$$

We then have

$$\begin{aligned} &\sum_{i,j=1}^n\xi_i\xi_j(\partial_{ij}f_0)(u)\\ =&\frac{1}{4}\cdot\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{3}{2}}|\xi\cdot\nabla A_k|^2B_k^{\frac{1}{2}} &&\text{(B.3)}\\ &+\frac{1}{4}\cdot\frac{1}{m}\sum_{k=1}^m B_k^{-\frac{3}{2}}|\xi\cdot\nabla B_k|^2A_k^{\frac{1}{2}} &&\text{(B.4)}\\ &-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}\langle\xi,(\nabla^2A_k)\xi\rangle B_k^{\frac{1}{2}} &&\text{(B.5)}\\ &-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^m A_k^{-\frac{1}{2}}B_k^{-\frac{1}{2}}(\nabla A_k\cdot\xi)(\nabla B_k\cdot\xi) &&\text{(B.6)}\\ &-\frac{1}{2}\cdot\frac{1}{m}\sum_{k=1}^m B_k^{-\frac{1}{2}}A_k^{\frac{1}{2}}\langle\xi,(\nabla^2B_k)\xi\rangle, &&\text{(B.7)} \end{aligned}$$

where

$$A_k=\beta|u|^2+(a_k\cdot u)^2+(a_k\cdot e_1)^2,\quad B_k=\beta|u|^2+2(a_k\cdot e_1)^2,$$

and we have denoted

$$\langle \xi, (\nabla^2 A_k)\xi \rangle = \sum_{i,j=1}^{n} \xi_i \xi_j \partial_{ij} A_k.$$

Thanks to the strong damping provided by A_k, it is tedious but not difficult to check that the terms (B.3), (B.5), (B.6) can be easily controlled with the help of Lemma B.3. The term (B.7) can be estimated in a similar way as in the estimate of (A.9) in the proof of Lemma A.10 (note that this is done in high probability therein!). The term (B.4) is also easy to handle. We omit further details. □

Acknowledgements

J. F. Cai was supported in part by Hong Kong Research Grant Council General Research Grant Nos. 16309518, 16309219, 16310620 and 16306821. Y. Wang was supported in part by the Hong Kong Research Grant Council General Research Grant Nos. 16306415 and 16308518.

References

[1] S. Bhojanapalli, N. Behnam, and N. Srebro, Global optimality of local search for low rank matrix recovery, Adv. Neural Infor. Proc. Syst., (2016), pp. 3873–3881.

[2] E. J. Candès and X. Li, Solving quadratic equations via PhaseLift when there are about as many equations as unknowns, Found. Comut. Math., 14(5) (2014), pp. 1017–1026.

[3] E. J. Candès, X. Li, and M. Soltanolkotabi, Phase retrieval via Wirtinger flow: Theory and algorithms, IEEE Trans. Inf. Theory, 61(4) (2015), pp. 1985–2007.

[4] E. J. Candès, T. Strohmer, and V. Voroninski, Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming, Commun. Pure Appl. Math., 66(8) (2013), pp. 1241–1274.

[5] J. Cai, M. Huang, D. Li and Y. Wang, Solving phase retrieval with random initial guess is nearly as good as by spectral initialization, Appl. Comput. Harmon. Anal., (2021).

[6] Y. Chen and E. J. Candès, Solving random quadratic systems of equations is nearly as easy as solving linear systems, Commun. Pure Appl. Math., 70(5) (2017), pp. 822–883.

[7] J. C. Dainty and J. R. Fienup, Phase retrieval and image reconstruction for astronomy, Image Recovery: Theory and Application, 231 (1987), 275.

[8] S. S. Du, C. Jin, J. D. Lee, and M. I. Jordan, Gradient descent can take exponential time to escape saddle points, Adv. Neural Infor. Proc. Syst., (2017), pp. 1067–1077.

[9] J. R. Fienup, Phase retrieval algorithms: a comparison, Appl. Opt., 21(15) (1982), pp. 2758–2769.

[10] B. Gao, X. Sun, Y. Wang, and Z. Xu, Perturbed amplitude flow for phase retrieval, IEEE Trans. Signal Process., 68 (2020), pp. 5427–5440.

[11] B. Gao and Z. Xu, Phaseless recovery using the Gauss–Newton method, IEEE Trans. Signal Process., 65(22) (2017), pp. 5885–5896.

[12] R. Ge, F. Huang, C. Jin, and Y. Yuan, Escaping from saddle points-online stochastic gradient for tensor decomposition, Conference on Learning Theory, (2015), pp. 797–842.

[13] R. Ge, J. Lee, C. Jin, and T. Ma, Matrix completion has no spurious local minimum, Adv. Neural Infor. Proc. Syst., (2016), pp. 2973–2981.

[14] R. W. Gerchberg, A practical algorithm for the determination of phase from image and diffraction plane pictures, Optik, 35 (1972), pp. 237–246.

[15] R. W. Gerchberg and W. O. Saxton, A practical algorithm for the determination of the phase from image and diffraction plane pictures, Optik, 35 (1972), pp. 237–246.

[16] R. W. Harrison, Phase problem in crystallography, Josa A, 10(5) (1993), pp. 1046–1055.

[17] M. Huang, M. J. Lai, A. Varghese, and Z. Xu, On DC based methods for phase retrieval, International Conference Approximation Theory, (2019), pp. 87–121.

[18] M. Huang and Y. Wang, Linear convergence of randomized Kaczmarz method for solving complex-valued phaseless equations, 2021, available: `http://arxiv.org/abs/2109.11811`.

[19] H. Jeong and C. S. Güntürk, Convergence of the randomized Kaczmarz method for phase retrieval, 2017, available: `http://arxiv.org/abs/1706.10291.`

[20] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, How to escape saddle points efficiently, Proceedings of the 34th International Conference on Machine Learning-Volume 70, (2017), pp. 1724–1732.

[21] C. Jin, P. Netrapalli, and M. I. Jordan, Accelerated gradient descent escapes saddle points faster than gradient descent, 2017, available: `http://arxiv.org/abs/1711.10456`.

[22] Z. Li, J. F. Cai, and K. Wei, Towards the optimal construction of a loss function without spurious local minima for solving quadratic equations, IEEE Trans. Inf. Theory, 66(5) (2020), pp. 3242–3260.

[23] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes, Annu. Rev. Phys. Chem., 59 (2008), pp. 387–410.

[24] R. P. Millane, Phase retrieval in crystallography and optics, J. Optical Soc. Am. A, 7(3) (1990), pp. 394–411.

[25] P. Netrapalli, P. Jain, and S. Sanghavi, Phase retrieval using alternating minimization, IEEE Trans. Signal Proc., 63(18) (2015), pp. 4814–4826.

[26] D. Park, A. Kyrillidis, and C. Caramanis, Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach, 2016, available:

`http://arxiv.org/abs/1609.03240`.

[27] H. Sahinoglou and S. D. Cabrera, On phase retrieval of finite-length sequences using the initial time sample, IEEE Trans. Circuits Syst., 38(8) (1991), pp. 954–958.

[28] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, Phase retrieval with application to optical imaging: a contemporary overview, IEEE Signal Process. Mag., 32(3) (2015), pp. 87–109.

[29] J. Sun, Q. Qu, and J, Wright, A geometric analysis of phase retrieval, Found. Comput. Math., 18(5) (2018), pp. 1131–1198.

[30] J. Sun, Q. Qu, and J, Wright, Complete dictionary recovery over the sphere I: Overview and the geometric picture, IEEE Trans. Inf. Theory, 63(2) (2016), pp. 853–884.

[31] Y. S. Tan and R. Vershynin, Phase retrieval via randomized kaczmarz: Theoretical guarantees, Information and Inference: A Journal IMA, 8(1) (2019), pp. 97–123.

[32] R. Vershynin, High-Dimensional Probability: An Introduction with Applications in Data Science, U.K.: Cambridge University Press, 2018.

[33] I. Waldspurger, A. d'Aspremont, and S. Mallat, Phase recovery, maxcut and complex semidefinite programming, Math. Prog., 149(1-2) (2015), pp. 47–81.

[34] A. Walther, The question of phase retrieval in optics, J. Mod. Opt., 10(1) (1963), pp. 41–49.

[35] G. Wang, G. B. Giannakis, and Y. C. Eldar, Solving systems of random quadratic equations via truncated amplitude flow, IEEE Trans. Inf. Theory, 64(2) (2018), pp. 773–794.

[36] K. Wei, Solving systems of phaseless equations via kaczmarz methods: a proof of concept study, Inverse Probl., 31(12) (2015), 125008.

[37] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms, The Journal of Machine Learning Research, 18(1) (2017), pp. 5164–5198.