Proceedings of Machine Learning Research vol 65:1–22, 2017

Correspondence retrieval

Alexandr Andoni ANDONI@CS.COLUMBIA.EDU
Daniel Hsu DJHSU@CS.COLUMBIA.EDU
Kevin Shi KSHI@CS.COLUMBIA.EDU
Computer Science Department, Columbia University

Xiaorui Sun XIAORUISUN@CS.COLUMBIA.EDU
Simons Institute for the Theory of Computing, UC Berkeley

Abstract

This article studies the correspondence retrieval problem: a set of k distinct but unknown points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d$ are to be recovered from the unordered collection of projection values $\langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle, \langle \boldsymbol{w}_i, \boldsymbol{x}_2 \rangle, \ldots, \langle \boldsymbol{w}_i, \boldsymbol{x}_k \rangle$ onto n known measurement vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$. Importantly, the correspondence of the k projections $\{\langle \boldsymbol{w}_i, \boldsymbol{x}_j \rangle\}_{j=1}^k$ across different measurements is unknown. A special case of this problem is the well-studied problem of (real-valued) phase retrieval. In the case of independent standard Gaussian measurement vectors, the main algorithm proposed in this work requires $n = d + 1$ measurements to correctly return the k unknown points with high probability. This number of measurements is optimal, and it is smaller than the number of measurements required for a stronger "for all" guarantee even in the phase retrieval setting. The algorithm is based on reductions to the Shortest Vector Problem on certain random lattices, and employs the Lenstra, Lenstra, & Lovász (1982) basis reduction algorithm in a manner similar to the Lagarias & Odlyzko (1985) algorithm for solving random instances of Subset Sum. Another algorithm, essentially due to Yi, Caramanis, & Sanghavi (2016), based on higher-order moments and tensor decompositions is shown to work in a setting where the projection values are corrupted by additive Gaussian noise, but it requires a significantly larger number of measurements.

1. Introduction

In (the real-variant of) the phase retrieval problem, an unknown vector $\boldsymbol{x} \in \mathbb{R}^d$ is to be recovered, up to sign, from magnitudes of projections $\left|\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle\right|$ onto n known measurement vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n \in \mathbb{R}^d$. The phase retrieval problem has a rich history in several engineering and scientific domains, especially when the $\boldsymbol{w}_i$ are Fourier basis vectors (see, e.g., Shechtman et al., 2015; Jaganathan et al., 2015, for an overview). The setting where the $\boldsymbol{w}_i$ are independent draws from certain probability distributions has been intensely studied in the past several years. Many algorithms based on numerical optimization (e.g., semidefinite programming, local optimization of convex and non-convex objectives) have been proven to solve the problem with high probability when provided enough measurements (Netrapalli et al., 2013; Candes et al., 2013; Candès and Li, 2014; Alexeev et al., 2014; Eldar and Mendelson, 2014; Candes et al., 2015a,b,c; Waldspurger et al., 2015; Chen and Candes, 2015; Sanghavi et al., 2016; Zhang and Liang, 2016; Wang et al., 2016; Kolte and Özgür, 2016; Gao and Xu, 2016; Sun et al., 2016).

In this paper, we consider a generalization of phase retrieval, which we call *correspondence retrieval*: a set of k distinct but unknown points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d$ are to be recovered from the unordered collection of projection values $\langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle, \langle \boldsymbol{w}_i, \boldsymbol{x}_2 \rangle, \ldots, \langle \boldsymbol{w}_i, \boldsymbol{x}_k \rangle$ onto n known measure-

ment vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$. Importantly, the correspondence of the k projections $\{\langle \boldsymbol{w}_i, \boldsymbol{x}_j \rangle\}_{j=1}^k$ across different measurements is unknown. Phase retrieval, as described above, is the special case where $k = 2$ and $\boldsymbol{x}_1 = -\boldsymbol{x}_2$; the two real numbers corresponding to each measurement are additive inverses ($\langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle$ and $\langle \boldsymbol{w}_i, \boldsymbol{x}_2 \rangle = -\langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle$).

We propose an algorithm for the case of independent standard Gaussian measurement vectors. For general k, the algorithm correctly recovers the unknown points with high probability from $n = d + 1$ measurements, assuming that the points are linearly independent. For the phase retrieval setting, a variant of the algorithm has the same guarantee without the linear independence assumption. Our algorithms are based on reductions to the Shortest Vector Problem (Ajtai, 1996) on certain random lattices; we prove that vectors provided by the Lenstra-Lenstra-Lovász basis reduction algorithm (henceforth LLL; Lenstra et al., 1982) yield to the correct solution for the correspondence retrieval problem. Our reduction generalizes an algorithm of Lagarias and Odlyzko (1985) for solving random instances of the Subset Sum Problem (Gary and Johnson, 1979, pg. 223). We note that Yi et al. (2014) establish the hardness of the phase retrieval via reduction *from* the Subset Sum Problem. Our algorithmic result can be viewed as a reduction in the other direction.

In the phase retrieval setting, our results show a gap between the number of measurement vectors required for all vectors $\boldsymbol{x} \in \mathbb{R}^d$ to be recoverable, and the number of random measurements sufficient for any particular vector to be recoverable. This is the same distinction between the "for all" and "for each" guarantees studied in the context of compressive sensing (Gilbert et al., 2007). Balan et al. (2006) prove that $n = 2d - 1$ measurement vectors are necessary for the "for all" guarantee, and also that the same number of typical measurement vectors are sufficient. Previous algorithmic results for phase retrieval require $n \geq Cd$ for some sufficiently large constant $C \geq 2$ or even $n \geq d \operatorname{poly} \log(d)$. Our algorithmic result has the "for each" guarantee: the $n = d + 1$ measurements suffice with high probability for the particular unknown vector of interest. Note that in the general correspondence retrieval problem, each measurement is comprised of k unordered real numbers, so the sufficiency of $d + 1$ measurements even when $k > 2$ is sensible.

We also describe an algorithm that works even when the measurements are corrupted by additive mean-zero Gaussian noise.[1] The algorithm is essentially the same as one proposed by Yi et al. (2016) for the related parameter estimation problem in the mixtures of linear regressions model; the main technique used is the method-of-moments and orthogonal tensor decomposition (Anandkumar et al., 2014). We observe that the moments used in the algorithm are invariant to the noise variance, and hence the algorithm is noise-robust in this sense. However, the number of measurements required by this algorithm, even when the noise is absent, is larger than that of the lattice-based algorithm. The moment-based algorithm appears to ignore consistency constraints across measurements that the lattice-based algorithm is able to exploit.

2. Setting and notations

This section describes the correspondence retrieval problem, notations and results concerning lattices and tensors, and the non-degeneracy condition required by the proposed algorithms.

1. In phase retrieval, noise is typically added to the square (magnitude) $\left|\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle\right|^2$ of the projections (Candes et al., 2015a, 2013). In our setting, independent noise is added to the k projections $\{\langle \boldsymbol{w}_i, \boldsymbol{x}_j \rangle\}_{j=1}^k$ themselves.

2.1. Correspondence retrieval problem

In an instance of the correspondence retrieval problem, k distinct but unknown points in $\mathbb{R}^d$, denoted by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \mathbb{R}^d$, are revealed through collections of noisy linear measurements.

The n measurement vectors, denoted by $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$, are i.i.d. random vectors in $\mathbb{R}^d$ with the standard multivariate Gaussian distribution $\mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{d\times d})$. For each $i \in \{1, 2, \ldots, n\}$, the i-th measurement is the unordered (multi-)set of k (Euclidean) inner products between $\boldsymbol{w}_i$ and the k points, corrupted by additive zero-mean Gaussian noise with variance σ^2:

$$\mathcal{M}_i^\sigma \;:=\; \{\langle \boldsymbol{w}_i, \boldsymbol{x}_1\rangle + \sigma\varepsilon_{i,1}, \langle \boldsymbol{w}_i, \boldsymbol{x}_2\rangle + \sigma\varepsilon_{i,2}, \ldots, \langle \boldsymbol{w}_i, \boldsymbol{x}_k\rangle + \sigma\varepsilon_{i,k}\}\,,$$

where the $\{\varepsilon_{i,j}\}_{1\le i\le n, 1\le j\le k}$ are i.i.d. $\mathrm{N}(0,1)$ random variables. The noiseless version of the problem has $\sigma^2 = 0$, and the measurements are denoted by $\mathcal{M}_i := \mathcal{M}_i^0$ for $i \in \{1, 2, \ldots, n\}$.

The goal is to (approximately) reconstruct the set of k unknown points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ (i.e., reconstruct up to reordering), from the data $(\boldsymbol{w}_1, \mathcal{M}_1^\sigma), (\boldsymbol{w}_2, \mathcal{M}_2^\sigma), \ldots, (\boldsymbol{w}_n, \mathcal{M}_n^\sigma)$.

2.2. Notations

The first m positive integers are denoted by $[m] := \{1, 2, \ldots, m\}$. The Euclidean inner product between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted by $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$, and the Euclidean norm is $\|\boldsymbol{v}\|_2 := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}$. The i-th largest singular value of a matrix $\boldsymbol{M}$ is denoted by $\sigma_i(\boldsymbol{M})$; the spectral norm (i.e., largest singular value) is also denoted by $\|\boldsymbol{M}\|_2$.

2.3. Lattices

An ordered basis $\boldsymbol{B} = [\boldsymbol{b}_1|\boldsymbol{b}_2|\cdots|\boldsymbol{b}_n] \in \mathbb{R}^{m\times n}$, arranged as columns in a rank n matrix, generates a lattice

$$\Lambda(\boldsymbol{B}) \;:=\; \left\{ \sum_{i=1}^n z_i \boldsymbol{b}_i : z_1, z_2, \ldots, z_r \in \mathbb{Z} \right\} \;\subset\; \mathbb{R}^m\,,$$

where $\mathbb{Z}$ denotes the set of integers. The Shortest Vector Problem is to find the shortest non-zero vector in the lattice:

$$\operatorname*{arg\,min}_{\boldsymbol{v}\in\Lambda(\boldsymbol{B})\setminus\{\boldsymbol{0}\}} \|\boldsymbol{v}\|_2\,.$$

The length of the shortest vector is denoted by $\lambda(\boldsymbol{B})$.

Current techniques for this problem involve "reducing" the input basis $\boldsymbol{B}$ so that it is at least somewhat well-conditioned in a certain sense. Lenstra et al. (1982) show that the first vector $\boldsymbol{b}_1$ in a suitably reduced basis $\boldsymbol{B}$ has length at most $2^{(n-1)/2}\cdot\lambda(\boldsymbol{B})$. They also give an algorithm (LLL) that, given a basis $\boldsymbol{B} \in \mathbb{Z}^{m\times n}$ with integer coefficients, computes a reduced basis $\boldsymbol{B}'$ with $\Lambda(\boldsymbol{B}') = \Lambda(\boldsymbol{B})$ in time polynomial in m, n, and $\log(\|\boldsymbol{B}\|_\infty)$, where $\|\boldsymbol{B}\|_\infty$ denotes the magnitude of the largest entry in $\boldsymbol{B}$. In this sense, LLL is a $2^{(n-1)/2}$-approximation algorithm for the Shortest Vector Problem.

An important concern with the use of LLL on bases with real-valued coefficients is numerical precision. There are two cases where precision needs to be considered: precision in the measurements, and precision in the internal arithmetic operations in LLL. We discuss these issues in Appendix B. To simplify the foregoing discussion, we assume that LLL may be run on input bases with real-valued coefficients.

2.4. Tensors

For a positive integer p, a real order-p tensor $\boldsymbol{T} \in \bigotimes_{i=1}^{p} \mathbb{R}^n$ is a p-linear function $\boldsymbol{T} \colon \mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$. We only require tensors of order two (i.e., matrices) and order three. The rank-one tensor $\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \cdots \otimes \boldsymbol{v}_p$, for vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in \mathbb{R}^n$, is the p-linear function satisfying

$$(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \cdots \otimes \boldsymbol{v}_p)(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p) \;=\; \langle \boldsymbol{v}_1, \boldsymbol{u}_1 \rangle \langle \boldsymbol{v}_2, \boldsymbol{u}_2 \rangle \cdots \langle \boldsymbol{v}_p, \boldsymbol{u}_p \rangle \,, \quad \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p \in \mathbb{R}^n \,.$$

We use the shorthand notation $\boldsymbol{v}^{\otimes p}$ for $\boldsymbol{v} \in \mathbb{R}^n$ to denote the (symmetric) rank-one tensor $\boldsymbol{v} \otimes \boldsymbol{v} \otimes \cdots \otimes \boldsymbol{v} \in \bigotimes_{j=1}^{p} \mathbb{R}^n$. For $p = 2$, this is the symmetric outer product of a vector: $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}^\top$. We may also identify a tensor $\boldsymbol{T} \in \bigotimes_{i=1}^{p} \mathbb{R}^n$ with a multi-index array of n^p real numbers; the $(i_1, i_2, \ldots, i_p)$-th entry is $\boldsymbol{T}(\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \ldots, \boldsymbol{e}_{i_p})$, where $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are the standard coordinate basis vectors for $\mathbb{R}^n$.

2.5. Non-degeneracy conditions

Arrange the k unknown points in the matrix $\boldsymbol{X} := [\boldsymbol{x}_1 | \boldsymbol{x}_2 | \cdots | \boldsymbol{x}_k] \in \mathbb{R}^{d \times k}$. Our main algorithms require $\boldsymbol{X}$ to have $\operatorname{rank}(\boldsymbol{X}) = k$—i.e., the points must be linearly independent.

We measure how ill-conditioned $\boldsymbol{X}$ is in two ways. The first is based on the singular values $\sigma_1(\boldsymbol{X}) \geq \sigma_2(\boldsymbol{X}) \geq \cdots \geq \sigma_k(\boldsymbol{X})$ of $\boldsymbol{X}$, primarily through the ratio $\kappa(\boldsymbol{X}) := \sigma_1(\boldsymbol{X}) / \sigma_k(\boldsymbol{X})$. The second is $\lambda(\boldsymbol{X})$, the length of the shortest non-zero vector in the lattice $\Lambda(\boldsymbol{X})$. The quantities $\kappa(\boldsymbol{X})$ and $\lambda(\boldsymbol{X})$ are related in the following proposition, which is proved in Appendix D.1.

Proposition 1 $\lambda(\boldsymbol{X}) \geq \min_{i \in [k]} \|\boldsymbol{x}_i\|_2 \cdot 2\kappa(\boldsymbol{X}) / (\kappa(\boldsymbol{X})^2 + 1)$.

For $k = 2$ (the phase retrieval setting), a variant of our lattice-based algorithm requires $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, but permits the points to be linearly dependent.

3. Noiseless correspondence retrieval

This section describes lattice-based algorithms for the noiseless correspondence retrieval problem.

3.1. Algorithm description

Our main algorithm, specified in Algorithm 1, is based on reductions to the Shortest Vector Problem in lattices. Using information from $d + 1$ measurements and the input parameter $\beta > 0$, the algorithm constructs k lattice bases with the following properties. First, for each $t \in [k]$, the only short vectors in the t-th lattice reveal which elements in the first d measurements correspond to the unknown vector $\boldsymbol{x}_t$. Second, when β is sufficiently large, all other vectors in the lattices are longer by exponentially-large factors. This lattice construction is based on the algorithm of Lagarias and Odlyzko (1985) for solving random instances of the Subset Sum Problem via reduction to the Shortest Vector Problem. Our algorithm similarly approximately solves these Shortest Vector Problem instances using LLL to obtain the correspondence information, and then recovers all of the k unknown points by solving systems of linear equations from the first d measurements.

3.2. Main result and analysis

The main performance guarantee for Algorithm 1 is given in Theorem 2 below.

Algorithm 1 Lattice-based algorithm for noiseless correspondence retrieval

input Data $(\boldsymbol{w}_i, \mathcal{M}_i)$ for $i \in [d+1]$, parameter $\beta > 0$.

output Set of points $\{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \dots, \hat{\boldsymbol{x}}_k\}$, or "failure".

1: **if** $\boldsymbol{W} := [\boldsymbol{w}_1 | \boldsymbol{w}_2 | \cdots | \boldsymbol{w}_d]^\top$ is singular **then**
2: **return** "failure".
3: **end if**
4: Let $y_{i,1}, y_{i,2}, \dots, y_{i,k}$ be an arbitrary ordering the elements of $\mathcal{M}_i$, for each $i \in [d+1]$.
5: Define $\boldsymbol{a} = (a_{i,j} : i \in [d], j \in [k]) \in \mathbb{R}^{dk}$ by

$$a_{i,j} \;:=\; \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle y_{i,j}\,,$$

where $\tilde{\boldsymbol{w}}_i$ is the i-th column of $\boldsymbol{W}^{-1}$.
6: **for** $t = 1, 2, \dots, k$ **do**
7: Construct basis

$$\boldsymbol{B}^{(t)} \;=\; \begin{bmatrix} \boldsymbol{b}_0^{(t)} & \boldsymbol{b}_{1,1}^{(t)} & \cdots & \boldsymbol{b}_{d,k}^{(t)} \end{bmatrix} \;:=\; \left[\begin{array}{c|c} \boldsymbol{I}_{dk+1} & \\ \hline \beta y_{d+1,t} & -\beta \boldsymbol{a}^\top \end{array}\right] \;\in\; \mathbb{R}^{(dk+2)\times(dk+1)}\,.$$

8: Let $L^{(t)}(\hat{z}_0, \hat{\boldsymbol{z}}) := \hat{z}_0 \boldsymbol{b}_0^{(t)} + \sum_{i,j} \hat{z}_{i,j} \boldsymbol{b}_{i,j}^{(t)} \in \Lambda(\boldsymbol{B}^{(t)})$ for $(\hat{z}_0, \hat{\boldsymbol{z}}) \in \mathbb{Z} \times \mathbb{Z}^{dk}$ be the vector returned by LLL as an approximate solution to Shortest Vector Problem for $\Lambda(\boldsymbol{B}^{(t)})$.
9: **if** the $(dk+2)$-th coordinate of $L^{(t)}(\hat{z}_0, \hat{\boldsymbol{z}})$ is non-zero **then**
10: **return** "failure".
11: **end if**
12: Let $\hat{\boldsymbol{x}}_t$ be a solution to the system of linear equations (in $\boldsymbol{x} \in \mathbb{R}^d$)

$$\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \;=\; y_{i,j}\,, \qquad (i,j) \in [d] \times [k] \,.\, \hat{z}_{i,j} \neq 0\,,$$

or $\mathbf{0}$ if no solution exists.
13: **end for**
14: **return** $\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \dots, \hat{\boldsymbol{x}}_k$.

Theorem 2 *Assume $\boldsymbol{X} = [\boldsymbol{x}_1 | \boldsymbol{x}_2 | \cdots | \boldsymbol{x}_k] \in \mathbb{R}^{d\times k}$ has $\operatorname{rank}(\boldsymbol{X}) = k$. For any $\delta \in (0,1)$, if*

$$\beta \;\geq\; \frac{16 \cdot \left(2 \cdot 2^{dk/2} \cdot \sqrt{d+1} + 1\right)^{dk+1} \cdot 2^{dk/2} \cdot d \cdot \sqrt{d+1} \cdot \left(2\sqrt{d} + \sqrt{2\ln(8/\delta)}\right) \cdot k^2}{\pi \cdot \delta^2 \cdot \lambda(\boldsymbol{X})}\,,$$

then with probability at least $1 - \delta$, Algorithm 1 returns $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k\}$.

Numerical issues and running time are discussed in Appendix B. The rest of this subsection is devoted to the proof of Theorem 2.

Let $R := 2^{dk/2} \cdot \sqrt{d+1}$, and let $\mathcal{Z}_R := \{(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^{dk} : 0 < z_0^2 + \|\boldsymbol{z}\|_2^2 \leq R^2\}$. For $\delta \in (0,1)$, define

$$r_\delta \;:=\; \sqrt{\frac{\pi}{2} \cdot \left(\frac{\delta}{k|\mathcal{Z}_R|}\right)^2 \cdot \frac{\lambda(\boldsymbol{X})^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2dk|\mathcal{Z}_R|}\right)^2}{\left(2\sqrt{d} + \sqrt{2\ln(4/\delta)}\right)^2}}\,.$$

The coefficient vectors in $\mathcal{Z}_R$ include all those that could potentially determine lattice vectors in $\Lambda(\boldsymbol{B}^{(t)})$ for $t \in [k]$ with length at most R. Below, we prove that these lattice vectors either provide the correspondence information needed to recover the unknown points (and have length $\ll R$), or they have length more than βr_δ with high probability. A crude bound on the cardinality of $\mathcal{Z}_R$ is

$$|\mathcal{Z}_R| \;\le\; \left|\{-\lfloor R\rfloor, -\lfloor R\rfloor + 1, \ldots, \lfloor R\rfloor - 1, \lfloor R\rfloor\}\right|^{dk+1} \;\le\; \left(2 \cdot 2^{(dk+1)/2} \cdot \sqrt{d+1} + 1\right)^{dk+1}.$$

For each $i \in [d]$, let $\pi_i \colon [k] \to [k]$ denote the (unknown) permutation on $[k]$ that determines the arbitrary ordering of $\mathcal{M}_i$ from Algorithm 1:

$$y_{i,j} \;=\; \langle \boldsymbol{w}_i, \boldsymbol{x}_{\pi_i(j)} \rangle\,, \quad i \in [d], j \in [k]\,.$$

Also, for $\delta \in (0,1)$, let $\mathcal{E}_\delta$ be the event that

1. the smallest singular value of $\boldsymbol{W}$ is bounded from below: $\sigma_d(\boldsymbol{W}) \ge \delta/(4\sqrt{d})$;
2. the spectral norm of $\boldsymbol{W}$ is bounded from above: $\|\boldsymbol{W}\|_2 \le 2\sqrt{d} + \sqrt{2\ln(4/\delta)}$;
3. for each $i \in [d]$, $j \in [k]$, and $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$ such that $|z_{i,j} - z_0| + \sum_{j' \ne j} |z_{i,j'}| > 0$,

$$\left\langle \boldsymbol{w}_i, (z_{i,j} - z_0)\boldsymbol{x}_{\pi_i(j)} + \sum_{j' \ne j} z_{i,j'} \boldsymbol{x}_{\pi_i(j')} \right\rangle^2 \;\ge\; \lambda(\boldsymbol{X})^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2dk|\mathcal{Z}_R|}\right)^2. \tag{1}$$

This event characterizes the properties needed from the first d measurements; Lemma 3 shows that it has large probability mass. The proof, given in Appendix D.2, is based on known properties of Gaussian random matrices.

Lemma 3 *For any $\delta \in (0,1)$, $\Pr(\mathcal{E}_\delta) \ge 1 - \delta$.*

We now show in Lemma 4 that, for each $t \in [k]$, there is a relatively short vector in $\Lambda(\boldsymbol{B}^{(t)})$ that provides the correspondence information needed to recover $\boldsymbol{x}_t$. We also show in Lemma 5 that when β is sufficiently large, other vectors in $\Lambda(\boldsymbol{B}^{(t)})$ are considerably longer, and hence cannot be returned by LLL.

To simplify notation, assume that $\pi_{d+1}(j) = j$ for each $j \in [k]$, so we have $y_{d+1,t} = \langle \boldsymbol{w}_{d+1}, \boldsymbol{x}_t \rangle$ for each $t \in [k]$. Using this notation, define $\boldsymbol{z}^{(t)} = (z_{i,j}^{(t)} : i \in [d], j \in [k]) \in \mathbb{Z}^{dk}$ for each $t \in [k]$ by

$$z_{i,j}^{(t)} \;:=\; \begin{cases} 1 & \text{if } \pi_i(j) = t\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Recall that for each $t \in [k]$, the lattice vector in $\Lambda(\boldsymbol{B}^{(t)})$ determined by coefficient vector $(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^{dk}$ is denoted by

$$L^{(t)}(z_0, \boldsymbol{z}) \;=\; z_0 \boldsymbol{b}_0^{(t)} + \sum_{i,j} z_{i,j} \boldsymbol{b}_{i,j}^{(t)}\,.$$

Observe that the coefficient vector $(z_0, \boldsymbol{z})$ is revealed in the first $dk+1$ coordinates of the lattice vector $L^{(t)}(z_0, \boldsymbol{z})$; the final coordinate of the lattice vector is used to make some vectors very long.

Lemma 4 *On the event $\mathcal{E}_\delta$, for each $t \in [k]$, $y_{d+1,t} = \langle \boldsymbol{w}_{d+1}, \boldsymbol{x}_t \rangle = \sum_{i,j} a_{i,j} z_{i,j}^{(t)}$. Also on this event, for each $t \in [k]$,*

$$L^{(t)}(1, \boldsymbol{z}^{(t)}) \;=\; \begin{bmatrix} 1 \\ \boldsymbol{z}^{(t)} \\ -\beta y_{d+1,t} + \beta \sum_{i,j} a_{i,j} z_{i,j}^{(t)} \end{bmatrix} \;=\; \begin{bmatrix} 1 \\ \boldsymbol{z}^{(t)} \\ 0 \end{bmatrix},$$
$$\left\| L^{(t)}(1, \boldsymbol{z}^{(t)}) \right\|_2 \;=\; \sqrt{d+1}\,.$$

Proof Assume $\mathcal{E}_\delta$ holds, which guarantees the existence of $\boldsymbol{W}^{-1}$ and thus permits the $\tilde{\boldsymbol{w}}_i$ to be well-defined. In this event, $\sum_{i=1}^d \tilde{\boldsymbol{w}}_i \boldsymbol{w}_i^\top = \boldsymbol{W}^{-1}\boldsymbol{W} = \boldsymbol{I}_d$. Therefore,

$$\begin{aligned} \sum_{i,j} a_{i,j} z_{i,j}^{(t)} \;&=\; \sum_{i=1}^d \sum_{j=1}^k \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle \langle \boldsymbol{w}_i, \boldsymbol{x}_{\pi_i(j)} \rangle z_{i,j}^{(t)} \\ &=\; \sum_{i=1}^d \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle \langle \boldsymbol{w}_i, \boldsymbol{x}_t \rangle \\ &=\; \boldsymbol{w}_{d+1}^\top \left(\sum_{i=1}^d \tilde{\boldsymbol{w}}_i \boldsymbol{w}_i^\top \right) \boldsymbol{x}_t \;=\; \langle \boldsymbol{w}_{d+1}, \boldsymbol{x}_t \rangle\,. \end{aligned}$$

The claim now follows by direct computation, using the above identity and the definition of $\boldsymbol{z}^{(t)}$. ■

Lemma 5 *For any $\delta \in (0,1)$, conditional on the event $\mathcal{E}_\delta$, with probability at least $1-\delta$ (over the choice of $\boldsymbol{w}_{d+1}$), for each $t \in [k]$, every coefficient vector $(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^{dk}$ that is not an integer multiple of $(1, \boldsymbol{z}^{(t)})$ satisfies*

$$\left\| L^{(t)}(z_0, \boldsymbol{z}) \right\|_2 \;>\; \min\left\{ R,\, \sqrt{z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_\delta^2} \right\}.$$

Proof Assume $\mathcal{E}_\delta$ holds. This implies, in particular, that $\boldsymbol{W}^{-1}$ and the $\tilde{\boldsymbol{w}}_i$ are well-defined. Fix $t \in [k]$, and let $\mathbb{Z}(1, \boldsymbol{z}^{(t)})$ denote the set of integer multiples of $(1, \boldsymbol{z}^{(t)})$. For any coefficient vector $(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^{dk}$,

$$\|L^{(t)}(z_0, \boldsymbol{z})\|_2^2 \;=\; \left\| z_0 \boldsymbol{b}_0^{(t)} + \sum_{i,j} z_{i,j} \boldsymbol{b}_{i,j}^{(t)} \right\|_2^2 \;=\; z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 \left(\sum_{i,j} a_{i,j} z_{i,j} - y_{d+1,t} z_0 \right)^2. \quad (2)$$

Observe that $\|L^{(t)}(z_0, \boldsymbol{z})\|_2 > R$ for all $(z_0, \boldsymbol{z}) \in (\mathbb{Z} \times \mathbb{Z}^{dk}) \setminus \mathcal{Z}_R$. Below, we prove that with probability at least $1 - \delta/k$, $\|L^{(t)}(z_0, \boldsymbol{z})\|_2^2 > z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_\delta^2$ for every $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R \setminus \mathbb{Z}(1, \boldsymbol{z}^{(t)})$. Combining this with a union bound over all choices of $t \in [k]$ proves the lemma.

Fix any such $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$, and consider the parenthesized term in Eq. (2) (without the squaring). By Lemma 4, the term expands to

$$\begin{aligned} \sum_{i,j} a_{i,j} z_{i,j} - y_{d+1,t} z_0 \;&=\; \sum_{i,j} a_{i,j} \left(z_{i,j} - z_{i,j}^{(t)} z_0 \right) \\ &=\; \sum_{i,j} \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle \langle \boldsymbol{w}_i, \boldsymbol{x}_{\pi_i(j)} \rangle \left(z_{i,j} - z_{i,j}^{(t)} z_0 \right) \;=\; \langle \boldsymbol{w}_{d+1}, \boldsymbol{v} \rangle\,, \end{aligned}$$

where

$$\boldsymbol{v} \;:=\; \sum_{i,j} \langle \boldsymbol{w}_i, \boldsymbol{x}_{\pi_i(j)} \rangle \left(z_{i,j} - z_{i,j}^{(t)} z_0 \right) \tilde{\boldsymbol{w}}_i \,.$$

Because $\boldsymbol{w}_{d+1} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, the final expression is a $\mathrm{N}(0, \|\boldsymbol{v}\|_2^2)$ random variable, and hence by Proposition 10 (given in Appendix A),

$$\Pr\left(\langle \boldsymbol{w}_{d+1}, \boldsymbol{v} \rangle^2 \leq \frac{\pi}{2} \cdot \left(\frac{\delta}{k|\mathcal{Z}_R|} \right)^2 \cdot \|\boldsymbol{v}\|_2^2 \right) \;\leq\; \frac{\delta}{k|\mathcal{Z}_R|} \,. \tag{3}$$

We show below that, on the event $\mathcal{E}_\delta$,

$$\|\boldsymbol{v}\|_2^2 \;\geq\; \frac{\lambda(\boldsymbol{X})^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2dk|\mathcal{Z}_R|} \right)^2}{\left(2\sqrt{d} + \sqrt{2\ln(2/\delta)} \right)^2} \,. \tag{4}$$

Using this bound with Eq. (3) and a union bound, it follows that with probability at least $1 - \delta/k$, we have $\|L^{(t)}(z_0, \boldsymbol{z})\|_2^2 > z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_\delta^2$ for all $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R \setminus \mathbb{Z}(1, \boldsymbol{z}^{(t)})$.

We now prove the bound in Eq. (4) on the event $\mathcal{E}_\delta$. Because the $\tilde{\boldsymbol{w}}_i$ are the columns of $\boldsymbol{W}^{-1}$, we may write $\boldsymbol{v} = \boldsymbol{W}^{-1}\boldsymbol{c}$ for $\boldsymbol{c} = (c_1, c_2, \ldots, c_d)$, where

$$c_i \;:=\; \sum_{j=1}^{k} \langle \boldsymbol{w}_i, \boldsymbol{x}_{\pi_i(j)} \rangle \left(z_{i,j} - z_{i,j}^{(t)} z_0 \right) \;=\; \left\langle \boldsymbol{w}_i, \left(z_{i,\pi_i^{-1}(t)} - z_0 \right) \boldsymbol{x}_t + \sum_{j \in [k] : \pi_i(j) \neq t} z_{i,j} \boldsymbol{x}_{\pi_i(j)} \right\rangle$$

for each $i \in [d]$. Therefore, $\|\boldsymbol{v}\|_2^2$ may be bounded below as

$$\|\boldsymbol{v}\|_2^2 \;\geq\; \sigma_d(\boldsymbol{W}^{-1})^2 \cdot \sum_{i=1}^{d} c_i^2 \;=\; \frac{1}{\|\boldsymbol{W}\|_2^2} \cdot \sum_{i=1}^{d} c_i^2 \,.$$

Since $(z_0, \boldsymbol{z}) \notin \mathbb{Z}(1, \boldsymbol{z}^{(t)})$, at least one of the following is true:

1. there exists $i \in [d]$ such that $z_{i,\pi_i^{-1}(t)} \neq z_0$;
2. there exists $i \in [d]$ and $j \in [k] \setminus \{\pi_i^{-1}(t)\}$ such that $z_{i,\pi_i(j)} \neq 0$.

In either case, there exists $i \in [d]$ such that $|z_{i,\pi_i^{-1}(t)} - z_0| + \sum_{j \in [k] : \pi_i(j) \neq t} |z_{i,\pi_i(j)}| > 0$, so using the third condition in the event $\mathcal{E}_\delta$,

$$\sum_{i=1}^{d} c_i^2 \;\geq\; \lambda(\boldsymbol{X})^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2dk|\mathcal{Z}_R|} \right)^2 .$$

Combining this with the upper-bound $\|\boldsymbol{W}\|_2 \leq 2\sqrt{d} + \sqrt{2\ln(2/\delta)}$ from the second condition in the event $\mathcal{E}_\delta$ proves the required lower-bound on $\|\boldsymbol{v}\|_2^2$ from Eq. (4). ■

We now prove Theorem 2. With probability at least $1 - \delta/2$ (over the choice of $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d$),

1. Event $\mathcal{E}_{\delta/2}$ holds (Lemma 3).

Moreover, conditional on $\mathcal{E}_{\delta/2}$,

2. $\|L^{(t)}(1, \boldsymbol{z}^{(t)})\|_2 = \sqrt{d+1}$ for each $t \in [k]$ (Lemma 4);

and, with probability at least $1 - \delta/2$ (over the choice of $\boldsymbol{w}_{d+1}$),

3. for each $t \in [k]$, every non-zero vector in $L^{(t)}(z_0, \boldsymbol{z}) \in \Lambda(\boldsymbol{B}^{(t)})$ for $(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^{dk}$ with length at most $R = 2^{(dk+1)/2}\sqrt{d+1}$ is either an integer multiple of $L^{(t)}(1, \boldsymbol{z}^{(t)})$, or has length $\|L^{(t)}(z_0, \boldsymbol{z})\|_2 > \sqrt{z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_{\delta/2}^2}$; the length in this latter case is more than R when $\beta \geq R/r_{\delta/2}$ (Lemma 5).

Statements 1–3 above hold together with probability at least $1 - \delta$, so we assume that they hold. In particular, Algorithm 1 does not return "failure" upon checking if $\boldsymbol{W}$ singular. As long as $\beta \geq R/r_{\delta/2}$, for each $t \in [k]$, the approximate solution returned by LLL for $\Lambda(\boldsymbol{B}^{(t)})$ is $L^{(t)}(\hat{z}_0, \hat{\boldsymbol{z}}) = L^{(t)}(c, c\boldsymbol{z}^{(t)})$ for some $c \neq 0$. The $(dk+2)$-th coordinate of this vector is zero—so Algorithm 1 does not return "failure" on account of this check—and $\hat{\boldsymbol{x}}_t$ is obtained as a solution to the system of linear equations

$$\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \;=\; y_{i,j}\,, \qquad (i,j) \in [d] \times [k] \,.\, cz_{i,j}^{(t)} \neq 0\,.$$

By the definition of $\boldsymbol{z}^{(t)}$ and non-singularity of $\boldsymbol{W}$, we have $\hat{\boldsymbol{x}}_t = \boldsymbol{x}_t$ for all $t \in [k]$. This completes the proof of Theorem 2.

3.3. Phase retrieval

The special case of correspondence retrieval where $k = 2$ and $\boldsymbol{x}_1 = -\boldsymbol{x}_2 \neq \boldsymbol{0}$ is known as the (real-valued) phase retrieval problem, as described in the introduction. Indeed, it is easy to see that the general $k = 2$ correspondence retrieval problem may be reduced to this case by "centering" the measurements. However, the unknown points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are no longer linearly independent, so Algorithm 1 is not directly applicable.

A simple fix is to pick a random vector $\boldsymbol{z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, and replace each unordered measurement $\mathcal{M}_i = \{\langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle, \langle \boldsymbol{w}_i, \boldsymbol{x}_2 \rangle\}$ with $\mathcal{M}_i' := \{\langle \boldsymbol{w}_i, \boldsymbol{z} \rangle + \langle \boldsymbol{w}_i, \boldsymbol{x}_1 \rangle, \langle \boldsymbol{w}_i, \boldsymbol{z} \rangle + \langle \boldsymbol{w}_i, \boldsymbol{x}_2 \rangle\}$. The points to recover become $\boldsymbol{z} + \boldsymbol{x}$ and $\boldsymbol{z} - \boldsymbol{x}$, where $\boldsymbol{x} := \boldsymbol{x}_1 = -\boldsymbol{x}_2$. Let $\tilde{\boldsymbol{X}} := [\boldsymbol{z} + \boldsymbol{x} | \boldsymbol{z} - \boldsymbol{x}] \in \mathbb{R}^{d \times 2}$. The following proposition gives a bound on $\kappa(\tilde{\boldsymbol{X}}) = \sigma_1(\tilde{\boldsymbol{X}})/\sigma_2(\tilde{\boldsymbol{X}})$; its proof is given in Appendix D.3.

Proposition 6 *For any vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, the matrix $\boldsymbol{M} := [\boldsymbol{a} + \boldsymbol{b} | \boldsymbol{a} - \boldsymbol{b}] \in \mathbb{R}^{d \times 2}$ satisfies*

$$\frac{\sigma_1(\boldsymbol{M})}{\sigma_2(\boldsymbol{M})} \;\leq\; \frac{r + 1/r}{\left|\sin(\theta)\right|}\,,$$

where $r := \|\boldsymbol{a}\|_2 / \|\boldsymbol{b}\|_2$, and θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

It is easy to see that

$$\kappa(\tilde{\boldsymbol{X}}) \;\leq\; \frac{r + 1/r}{\left|\sin(\theta)\right|} \;\leq\; O\left(\frac{\|\boldsymbol{x}\|_2}{\sqrt{d}} + \frac{\sqrt{d}}{\|\boldsymbol{x}\|_2}\right)$$

with high probability, and hence Algorithm 1 may be applied.

We can also give a direct algorithm for solving the phase retrieval problem via LLL, with qualitatively the same guarantees as Algorithm 1, where $\|\boldsymbol{x}\|_2$ replaces the role of $\lambda(\boldsymbol{X})$. The details are given in Appendix C.

Number of measurements. Our algorithms require $n = d + 1$ measurements for exact recovery, which is the best possible (in dimension $d \geq 2$), even in this phase retrieval setting. With only d linearly independent measurement vectors, no algorithm can distinguish among 2^{d-1} distinct solutions (of the form $\{\boldsymbol{W}^{-1}\operatorname{diag}(\boldsymbol{s})\boldsymbol{W}\boldsymbol{x}, -\boldsymbol{W}^{-1}\operatorname{diag}(\boldsymbol{s})\boldsymbol{W}\boldsymbol{x}\}$ for $\boldsymbol{s} \in \{\pm 1\}^d$) that give rise to the same d measurements.

As discussed in the introduction, Balan et al. (2006) prove that $n = 2d-1$ measurement vectors (whether random or deterministic) are necessary to ensure that every non-zero $\boldsymbol{x} \in \mathbb{R}^d$ can be recovered, up to sign, from measurements with these measurement vectors. Because our algorithms only use $d+1$ (Gaussian) measurement vectors, they must be insufficient for recovering some $\boldsymbol{x}$ up to sign (in dimension $d \geq 3$), even though for any fixed $\boldsymbol{x}$, they suffice with high probability.

4. Noisy correspondence retrieval

This section sketches a moment-based algorithm for the noisy correspondence retrieval problem.

4.1. Main idea

The algorithm is based on decomposing the following moments involving the k unknown points:

$$\boldsymbol{M} := \sum_{j=1}^{k} \boldsymbol{x}_j^{\otimes 2} \in \mathbb{R}^{d\times d} \qquad \text{and} \qquad \boldsymbol{T} := \sum_{j=1}^{k} \boldsymbol{x}_j^{\otimes 3} \in \mathbb{R}^{d\times d\times d}.$$

Under the condition $\operatorname{rank}(\boldsymbol{X}) = k$, there is an efficient algorithm based on tensor decompositions that, if given $\boldsymbol{M}$ and $\boldsymbol{T}$ up to some sufficiently small error as inputs, returns accurate estimates of the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ up to reordering (see, e.g., Anandkumar et al., 2014).

The crucial idea is that the moment matrix $\boldsymbol{M}$ and tensor $\boldsymbol{T}$ can be estimated from the data $(\boldsymbol{w}_1, \mathcal{M}_1^\sigma), (\boldsymbol{w}_2, \mathcal{M}_2^\sigma), \ldots, (\boldsymbol{w}_n, \mathcal{M}_n^\sigma)$, even though the measurements are unordered. This was observed by Yi et al. (2016) in the case of a related (and indeed, more difficult) model of mixtures of linear regressions. In their model, there is no noise (i.e., $\sigma = 0$), but instead of observing all of $\mathcal{M}_i$, only a random element of $\mathcal{M}_i$ is observed (and this random choice is independent of the random measurement vectors, and identically distributed across all n measurements). Yi et al. give an algorithm for learning the k unknown points when n is sufficiently large (nearly linear in d, polynomial in k and $\kappa(\boldsymbol{X})$).[2] Therefore, it is clear that the noiseless correspondence retrieval problem may be reduced to their noiseless mixtures of linear regressions problem.

Our main observation is that the same estimators designed for the noiseless setting may also be applied in the noisy setting.

2. Yi et al. (2016) also give a hybrid algorithm that combines alternating minimization with the moment-based algorithm. This hybrid algorithm can exactly recover the k unknown points in the noiseless setting.

4.2. Moment estimators

To estimate $\boldsymbol{M}$ and $\boldsymbol{T}$, we use

$$\widehat{\boldsymbol{M}} := \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{2}\sum_{j=1}^{k}\left(\langle \boldsymbol{w}_i, \boldsymbol{x}_j\rangle + \sigma\varepsilon_{i,j}\right)^2\left(\boldsymbol{w}_i^{\otimes 2} - \boldsymbol{I}_d\right)\right\}$$

$$\text{and}\quad \widehat{\boldsymbol{T}} := \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{6}\sum_{j=1}^{k}\left(\langle \boldsymbol{w}_i, \boldsymbol{x}_j\rangle + \sigma\varepsilon_{i,j}\right)^3\left(\boldsymbol{w}_i^{\otimes 3} - \mathcal{T}(\boldsymbol{w}_i)\right)\right\},$$

respectively. Here, for any vector $\boldsymbol{v} \in \mathbb{R}^d$, the third-order tensor $\mathcal{T}(\boldsymbol{v})$ is defined by $\mathcal{T}(\boldsymbol{v}) := \sum_{j=1}^{d}(\boldsymbol{v}\otimes\boldsymbol{e}_j\otimes\boldsymbol{e}_j + \boldsymbol{e}_j\otimes\boldsymbol{v}\otimes\boldsymbol{e}_j + \boldsymbol{e}_j\otimes\boldsymbol{e}_j\otimes\boldsymbol{v})$, where $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d$ is any fixed orthonormal basis for $\mathbb{R}^d$. The i-th term in each of $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{T}}$ is symmetric with respect to the k values in $\mathcal{M}_i^\sigma$, and hence can be formed using just the unordered measurements.

The unbiasedness of $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{T}}$ in the noiseless case ($\sigma = 0$) follows immediately from the following proposition. We give a simple proof in Appendix D.4 for completeness.

Proposition 7 (Yi et al., 2016) *Let $\boldsymbol{w} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_d)$. For any vector $\boldsymbol{u} \in \mathbb{R}^d$,*

$$\mathbb{E}\left[\frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{u}\rangle^2\left(\boldsymbol{w}^{\otimes 2} - \boldsymbol{I}_d\right)\right] = \boldsymbol{u}^{\otimes 2}, \qquad \mathbb{E}\left[\frac{1}{6}\langle \boldsymbol{w}, \boldsymbol{u}\rangle^3\left(\boldsymbol{w}^{\otimes 3} - \mathcal{T}(\boldsymbol{w})\right)\right] = \boldsymbol{u}^{\otimes 3}.$$

In the noisy case, we have the following analogous proposition, which implies the unbiasedness of $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{T}}$ for any noise level $\sigma \geq 0$.

Proposition 8 *Let $\boldsymbol{w} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ and $\varepsilon \sim \mathrm{N}(0, 1)$ be independent. For any vector $\boldsymbol{u} \in \mathbb{R}^d$ and any $\sigma \geq 0$,*

$$\mathbb{E}\left[\frac{1}{2}\left(\langle \boldsymbol{w}, \boldsymbol{u}\rangle + \sigma\varepsilon\right)^2\left(\boldsymbol{w}^{\otimes 2} - \boldsymbol{I}_d\right)\right] = \boldsymbol{u}^{\otimes 2}, \quad \mathbb{E}\left[\frac{1}{6}\left(\langle \boldsymbol{w}, \boldsymbol{u}\rangle + \sigma\varepsilon\right)^3\left(\boldsymbol{w}^{\otimes 3} - \mathcal{T}(\boldsymbol{w})\right)\right] = \boldsymbol{u}^{\otimes 3}.$$

Proof This follows from Proposition 7 by replacing $\boldsymbol{w}$ and $\boldsymbol{u}$, respectively, with $\tilde{\boldsymbol{w}} := (\boldsymbol{w}, \varepsilon) \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{d+1})$ and $\tilde{\boldsymbol{u}} := (\boldsymbol{u}, \sigma) \in \mathbb{R}^{d+1}$; and considering the appropriate sub-matrix and sub-tensor. ■

Proposition 8 justifies the use of essentially the same moment-based algorithm of Yi et al. for the noisy correspondence retrieval problem:

1. Compute the estimates $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{T}}$ from $(\boldsymbol{w}_1, \mathcal{M}_1^\sigma), (\boldsymbol{w}_2, \mathcal{M}_2^\sigma), \ldots, (\boldsymbol{w}_n, \mathcal{M}_n^\sigma)$.
2. Apply the tensor decomposition algorithm of Anandkumar et al. (2014), and return the vectors from the approximate decomposition $\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \ldots, \hat{\boldsymbol{x}}_k$.

The analysis of Yi et al. can be used to give a bound on the number of measurements needed to accurately estimate the k unknown points: assuming $\max_{j\in k]}\|\boldsymbol{x}_j\|_2 = 1$, for any $\varepsilon, \delta \in (0, 1)$, if the number of measurements n satisfies

$$n \geq \tilde{O}\left(d \cdot \mathrm{poly}\left(\frac{k}{\sigma_k(\boldsymbol{X}_\sigma)}\right) \cdot \frac{\log(1/\delta)}{\varepsilon^2} + \frac{k^2}{\delta}\right),$$

then the algorithm returns $\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \ldots, \hat{\boldsymbol{x}}_k \in \mathbb{R}^d$ satisfying

$$\min_{\pi} \max_{j \in [k]} \|\hat{\boldsymbol{x}}_{\pi(j)} - \boldsymbol{x}_j\|_2 \;\leq\; \varepsilon\,,$$

with probability at least $1 - \delta$, where the min is over permutations $\pi \colon [k] \to [k]$. Here, the $\tilde{O}(\cdot)$ hides factors that are poly-logarithmic in those that appear, and $\boldsymbol{X}_\sigma$ is the $(d+1) \times k$ matrix that appends a row to $\boldsymbol{X}$ with all entries equal to σ. We omit a detailed bound and analysis because they are based entirely on the results of Yi et al., and the result is not comparable to the results we obtain in the noiseless setting with the lattice-based algorithms.

5. Discussion

The moment-based algorithm for the correspondence retrieval problem does not appear to efficiently use the information contained in individual measurements. By averaging over the measurements in the computation of $\widehat{\boldsymbol{M}}$ and $\widehat{\boldsymbol{T}}$, critical constraint information is lost. In contrast, the lattice-based algorithm does not average over the projection values nor the measurements themselves. It would be interesting to understand if there is indeed a gap between these distinct types of algorithms.

It would also be interesting to consider other classes of measurement vectors. Assuming a Gaussian distribution is convenient for analysis of our lattice-based algorithm, although it is plausible that other distributions satisfying some kind of anti-concentration condition at every point would also suffice. Handling certain discrete distributions would also simplify the numerical precision issues. The moment-based algorithm, however, critically relies on higher-order moment calculations specific to the Gaussian distribution. It is not clear to what extent that algorithm would work with other classes of measurement vectors. A plausible alternative is to use semidefinite programming to recover $\boldsymbol{M}$ and $\boldsymbol{T}$ (or other related moment tensors). Indeed, the results of Kueng et al. (2017) imply that $\boldsymbol{M}$ can be recovered from $O(dk)$ measurements, where the distribution of the measurement vectors may be Gaussian or from a certain class of finitely-supported distributions.

Our lattice-based algorithm cannot handle measurement noise, with the cryptographic hardness of the Shortest Vector Problem being the main barrier. There is also cryptographic evidence that even deterministic measurement errors make related problems computationally intractable (Alwen et al., 2013). In practice, LLL has been observed to find the shortest vector in lattices in low dimensions, and in high dimensions, its empirical performance is somewhat better than the worst-case approximation factor (Stehlé, 2010). Nevertheless, it is desirable to find different algorithms for phase retrieval and correspondence retrieval that do not use LLL but still work with the same optimal number of measurements.

Acknowledgments

AA was supported in part by a grant from the Simons Foundation (#491119 to Alexandr Andoni), a Google Faculty Research Award, and NSF award CCF-1617955. DH was supported in part by NSF awards DMR-1534910 and IIS-1563785, a Bloomberg Data Science Research Grant, and a Sloan Research Fellowship. KS was supported in part by NSF awards CCF-1423306, CNS-1552932, and DMR-1534910. XS was supported in part by NSF awards CCF-1149257 and CCF-1423100, as well as a grant from the Simons Foundation (#320173 to Xiaorui Sun). This work was done in part while DH and KS were research visitors and XS was a research fellow at the Simons Institute for the Theory of Computing.

References

Miklós Ajtai. Generating hard instances of lattice problems. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 99–108. ACM, 1996.

Boris Alexeev, Afonso S Bandeira, Matthew Fickus, and Dustin G Mixon. Phase retrieval with polarization. *SIAM Journal on Imaging Sciences*, 7(1):35–66, 2014.

Joel Alwen, Stephan Krenn, Krzysztof Pietrzak, and Daniel Wichs. Learning with rounding, revisited: New reduction, properties and applications. Cryptology ePrint Archive, Report 2013/098, 2013.

Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15 (1):2773–2832, 2014.

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

Emmanuel J Candès and Xiaodong Li. Solving quadratic equations via phaselift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014.

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

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

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

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015c.

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

Kenneth R Davidson and Stanislaw J Szarek. Local operator theory, random matrices and banach spaces. *Handbook of the geometry of Banach spaces*, 1(317-366):131, 2001.

Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM Journal on Matrix Analysis and Applications*, 9(4):543–560, 1988.

Yonina C Eldar and Shahar Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*, 36(3):473–494, 2014.

Bing Gao and Zhiqiang Xu. Gauss-newton method for phase retrieval. *arXiv preprint*, 2016.

Michael R Gary and David S Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. WH Freeman and Company, New York, 1979.

Anna C Gilbert, Martin J Strauss, Joel A Tropp, and Roman Vershynin. One sketch for all: fast algorithms for compressed sensing. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 237–246. ACM, 2007.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 1985.

Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. Phase retrieval: An overview of recent developments. *arXiv preprint arXiv:1510.07713*, 2015.

Ritesh Kolte and Ayfer Özgür. Phase retrieval via incremental truncated wirtinger flow. *arXiv preprint arXiv:1606.03196*, 2016.

Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

Jeffrey C Lagarias and Andrew M Odlyzko. Solving low-density subset sum problems. *Journal of the ACM*, 32(1):229–246, 1985.

Arjen Klaas Lenstra, Hendrik Willem Lenstra, and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

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

Sujay Sanghavi, Rachel Ward, and Chris D White. The local convexity of solving systems of quadratic equations. *Results in Mathematics*, pages 1–40, 2016.

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

Damien Stehlé. *Floating-Point LLL: Theoretical and Practical Aspects*, pages 179–213. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-02295-1. doi: 10.1007/978-3-642-02295-1_5.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 2379–2383. IEEE, 2016.

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

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

Xinyang Yi, Constantine Caramanis, and Sujay Sanghavi. Alternating minimization for mixed linear regression. In *ICML*, pages 613–621, 2014.

Xinyang Yi, Constantine Caramanis, and Sujay Sanghavi. Solving a mixture of many random linear equations by tensor decomposition and alternating minimization. *arXiv preprint arXiv:1608.05749*, 2016.

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

Appendix A. Gaussian inequalities

Theorem 9 (Edelman, 1988; Davidson and Szarek, 2001) *Let $\boldsymbol{Z}$ be an $n \times n$ matrix whose entries are i.i.d.* $\mathrm{N}(0,1)$ *random variables. For any $\eta \in (0,1)$,*

$$\Pr\left(\sigma_n(\boldsymbol{Z}) \leq \frac{\eta}{\sqrt{n}}\right) \;\leq\; \eta\,,$$

and

$$\Pr\left(\sigma_1(\boldsymbol{Z}) \geq 2\sqrt{d} + \sqrt{2\ln(1/\eta)}\right) \;\leq\; \eta\,.$$

The following proposition is based on elementary properties of the Gaussian distribution.

Proposition 10 *Let $Z \sim \mathrm{N}(0,1)$. For any $\eta \in (0,1)$, $\Pr(Z^2 \leq \pi\eta^2/2) \leq \eta$, and $\Pr(|Z| > \sqrt{2\ln(2/\eta)}) \leq \eta$.*

Proof The first bound is a standard Gaussian anti-concentration bound:

$$\Pr\left(|Z| \leq \sqrt{\frac{\pi}{2}}\eta\right) \;=\; \int_{-\sqrt{\frac{\pi}{2}}\eta}^{\sqrt{\frac{\pi}{2}}\eta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z \;\leq\; \int_{-\sqrt{\frac{\pi}{2}}\eta}^{\sqrt{\frac{\pi}{2}}\eta} \frac{1}{\sqrt{2\pi}}\,\mathrm{d}z \;=\; \eta\,.$$

The second bound is a standard upper-bound on the Gaussian tail. ■

Appendix B. Numerical issues

In this section, we discuss the numerical issues with Algorithm 1. We assume that the coefficients of the measurement vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d$ and the k unknown points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ are represented with sufficiently fine precision—say, with B bits of precision—and that the projection values $\langle \boldsymbol{w}_i, \boldsymbol{x}_j \rangle$ in the measurements are exact. Here, B should be large enough so that the Gaussian anti-concentration properties in the proof of Theorem 2 still hold (say, within a constant multiplicative factor). The anti-concentration property from Proposition 10 is used with η no smaller than $\lambda(\boldsymbol{X}) \cdot 2^{-\operatorname{poly}(d,k,\log(1/\delta))}$, so the number of bits needed is $\operatorname{poly}(d,k,\log(1/\delta))$ plus the number of bits needed to represent $\lambda(\boldsymbol{X})$, the length of the shortest vector in $\Lambda(\boldsymbol{X})$. Recall that $\lambda(\boldsymbol{X})$ is no larger than $\min_{j\in[k]} \|\boldsymbol{x}_j\|_2$ and, by Proposition 1, no smaller than $\min_{j\in[k]} \|\boldsymbol{x}_j\|_2 / \kappa(\boldsymbol{X})$.

The numerical work performed by Algorithm 1 is dominated by the calls to LLL and the solving of linear systems. Lemma 11 bounds how much smaller or larger the coefficients of the lattice basis (used in the calls to LLL) are relative to the projection values. Lemma 11 also bounds the condition number of the matrix involved in the linear system that is used to solve for the unknown points.

Lemma 11 *With probability at least* $1 - \delta$,

$$\begin{aligned}
\sigma_1(\boldsymbol{W}) &\le 2\sqrt{d} + \sqrt{2\ln(4/\delta)}\,, \\
\sigma_d(\boldsymbol{W}) &\ge \frac{\delta}{4\sqrt{d}}\,, \\
\beta \cdot |a_{i,j}| &\in \left[\frac{\beta \cdot \sqrt{\pi}\delta \cdot |y_{i,j}|}{4\sqrt{2}d\left(2\sqrt{d} + \sqrt{2\ln(4/\delta)}\right)}, \frac{\beta \cdot 4\sqrt{2d\ln(8d/\delta)} \cdot |y_{i,j}|}{\delta}\right], \quad i \in [d], j \in [k]\,.
\end{aligned}$$

Proof From Theorem 9, it follows that

$$\Pr\left(\sigma_1(\boldsymbol{W}) \le 2\sqrt{d} + \sqrt{2\ln(4/\delta)} \quad \text{and} \quad \sigma_d(\boldsymbol{W}) \ge \delta/(4\sqrt{d})\right) \ge 1 - \frac{\delta}{2}\,.$$

Condition on this $1 - \delta/2$ probability event. Recall that $\tilde{\boldsymbol{w}}_i$ is the i-th column of $\boldsymbol{W}^{-1}$. The distribution of each $\langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle$ is $\mathrm{N}(0, \|\tilde{\boldsymbol{w}}_i\|_2^2)$, and

$$\frac{1}{\sigma_1(\boldsymbol{W})} \le \|\tilde{\boldsymbol{w}}_i\|_2 \le \frac{1}{\sigma_d(\boldsymbol{W})}\,.$$

So, by Proposition 10 and union bound,

$$\Pr\left(\forall i \in [d]\,.\, \frac{\sqrt{\pi}\delta}{4\sqrt{2}d\sigma_1(\boldsymbol{W})} \le |\langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle| \le \frac{\sqrt{2\ln(8d/\delta)}}{\sigma_d(\boldsymbol{W})}\right) \ge 1 - \frac{\delta}{2}\,.$$

Since $a_{i,j} = \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle y_{i,j}$, combining these probability bounds proves the claim. ■

When calling LLL, we may treat the lattice basis coefficients as integers by rescaling. By Lemma 11, the number of bits required to represent these coefficients may grow from B to

$$B + O\left(\log \max\left\{\frac{d^{3/2} + d\sqrt{\log(1/\delta)}}{\beta\delta}, \frac{\beta\sqrt{d\log(d/\delta)}}{\delta}\right\}\right)\,.$$

With the required value of β from the statement of Theorem 2, the running time of LLL—and also of Algorithm 1—is therefore $\mathrm{poly}(d, k, \log(B), \log(\kappa(\boldsymbol{X})), \log(1/\delta))$.

Appendix C. Direct algorithm for phase retrieval

In the phase retrieval problem, there is a single hidden vector $\boldsymbol{x}$, and for each $i \in [d+1]$, we draw $\boldsymbol{w}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ and observe $y_i := \left|\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle\right|$. Our goal is to recover $\boldsymbol{x}$ by finding the vector of unknown signs $\boldsymbol{s} := (s_1, s_2, \ldots, s_d) \in \{\pm 1\}^d$, where $s_i := \mathrm{sign}(\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle)$ for each $i \in [d]$. A modified version of our main algorithm, specified in Algorithm 2, constructs a lattice where the shortest vector's coefficients are exactly the same as $\boldsymbol{s}$ or $-\boldsymbol{s}$.

The performance guarantee of this algorithm is given below in an analogous result to Theorem 2.

Algorithm 2 Lattice-based algorithm for phase retrieval

input Data $(\boldsymbol{w}_i, y_i)$ for $i \in [d+1]$, parameter $\beta > 0$.

output Hidden point $\hat{\boldsymbol{x}}$ (up to a sign), or "failure".

1: **if** $\boldsymbol{W} := [\boldsymbol{w}_1|\boldsymbol{w}_2|\cdots|\boldsymbol{w}_d]^\top$ is singular **then**

2: **return** "failure".

3: **end if**

4: Define $\boldsymbol{a} = (a_i : i \in [d]) \in \mathbb{R}^d$ by

$$a_i \;:=\; \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i \rangle y_i \,,$$

where $\tilde{\boldsymbol{w}}_i$ is the i-th column of $\boldsymbol{W}^{-1}$.

5: Construct basis

$$\boldsymbol{B} \;=\; \begin{bmatrix} \boldsymbol{b}_0 & \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_d \end{bmatrix} \;:=\; \left[\begin{array}{c|c} \boldsymbol{I}_{d+1} & \\ \hline \beta y_{d+1} & -\beta \boldsymbol{a}^\top \end{array}\right] \;\in\; \mathbb{R}^{(d+2)\times(d+1)} \,.$$

6: Let $(\hat{z}_0, \hat{\boldsymbol{z}}) \in \mathbb{Z} \times \mathbb{Z}^d$ specify an approximate solution $\hat{z}_0 \boldsymbol{b}_0 + \sum_i \hat{z}_i \boldsymbol{b}_i \in \Lambda(\boldsymbol{B})$ to the Shortest Vector Problem for $\Lambda(\boldsymbol{B})$ using LLL.

7: **if** $|\hat{z}_0| = |\hat{z}_1| = |\hat{z}_2| = \cdots = |\hat{z}_d|$ is not true **then**

8: **return** "failure"

9: **end if**

10: Let $\hat{\boldsymbol{x}}$ be a solution to the system of linear equations (in $\boldsymbol{t} \in \mathbb{R}^d$)

$$\langle \boldsymbol{w}_i, \boldsymbol{t} \rangle \;=\; \operatorname{sign}(\hat{z}_i) y_i \,, \qquad i \in [d] \,.$$

11: **return** $\hat{\boldsymbol{x}}$.

Theorem 12 *For any $\delta \in (0,1)$, if*

$$\beta \;\geq\; \frac{2^{d/2}\sqrt{d+1} \cdot 2d \left(2\sqrt{d} + \sqrt{2\ln(4/\delta)}\right) \cdot \left(2 \cdot 2^{d/2}\sqrt{d+1} + 1\right)^{d+1}}{\delta^2 \|\boldsymbol{x}\|_2 \, \pi}$$

then with probability at least $1-\delta$, Algorithm 2 returns $\hat{\boldsymbol{x}} = \boldsymbol{x}$.

Let $\mathcal{E}_\delta$ be the event that

1. the smallest singular value of $\boldsymbol{W}$ is bounded from below: $\sigma_d(\boldsymbol{W}) \geq \delta/(4\sqrt{d})$;
2. the spectral norm of $\boldsymbol{W}$ is bounded from above: $\|\boldsymbol{W}\|_2 \leq 2\sqrt{d} + \sqrt{2\ln(4/\delta)}$;
3. for each $i \in [d]$ and $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$ such that $|z_i - z_0| > 0$,

$$\left\langle \boldsymbol{w}_i, (z_i - z_0)\boldsymbol{x} \right\rangle^2 \;\geq\; \|\boldsymbol{x}\|_2^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2d|\mathcal{Z}_R|}\right)^2 . \tag{5}$$

Also, let $R := 2^{d/2}\sqrt{d+1}$ and $\mathcal{Z}_R := \{(z_0, \boldsymbol{z}) \in \mathbb{Z} \times \mathbb{Z}^d : 0 < z_0^2 + \|\boldsymbol{z}\|_2^2 \leq R^2\}$.

Lemma 13 *For any $\delta \in (0,1)$, $\Pr(\mathcal{E}_\delta) \geq 1-\delta$.*

The proof of this lemma is completely analogous to that of Lemma 3, so we omit it.

Let $s_0 := \operatorname{sign}(\langle \boldsymbol{w}_{d+1}, \boldsymbol{x}\rangle)$. The following lemma shows there exists a short lattice vector which solves the recovery problem. Its proof is analogous to that of Lemma 4, so again we omit it.

Lemma 14 *On the event $\mathcal{E}_\delta$*

$$\begin{aligned} L(s_0, \boldsymbol{s}) &= \begin{bmatrix} 1 \\ \boldsymbol{s} \\ -\beta s_0 y_{d+1} + \beta \sum_i^d a_i s_i \end{bmatrix} = \begin{bmatrix} 1 \\ \boldsymbol{s} \\ 0 \end{bmatrix}, \\ \|L(s_0, \boldsymbol{s})\|_2 &= \sqrt{d+1}\,. \end{aligned}$$

Finally, we state a lemma that lower-bounds the length of lattice vectors that are not integer multiples of $L(s_0, \boldsymbol{s})$.

Lemma 15 *For any $\delta \in (0,1)$, conditioned on the event $\mathcal{E}_\delta$, for every coefficient vector $(z_0, \boldsymbol{z})$ that is not an integer multiple of $(s_0, \boldsymbol{s})$, we have*

$$\|L(z_0, \boldsymbol{z})\|_2^2 \;>\; \min\left\{R^2,\; z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_\delta^2\right\}.$$

where

$$r_\delta \;:=\; \delta^2 \cdot \frac{\|\boldsymbol{x}\|_2\, \pi}{2d|\mathcal{Z}_R|(2\sqrt{d} + \sqrt{2\ln(4/\delta)})}\,.$$

Proof Let $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$ be any coefficient vector. Then the last coordinate of the corresponding lattice vector is

$$\sum_{i=1}^d a_i z_i - z_0 y_{d+1} \;=\; \sum_{i=1}^d a_i z_i - z_0 s_0 \sum_{i=1}^d a_i s_i$$

(using the relation from Lemma 14 and the fact that $s_0^2 = 1$)

$$\begin{aligned} &= \sum_{i=1}^d a_i \left(z_i - z_0 s_0 s_i\right) \\ &= \sum_{i=1}^d \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i\rangle \left|\langle \boldsymbol{w}_i, \boldsymbol{x}\rangle\right| \left(z_i - z_0 s_0 s_i\right). \end{aligned}$$

Because $\boldsymbol{z}$ is not an integer multiple of $\boldsymbol{s}$ and $z_0 s_0$ is an integer, there exists an index $i^* \in [d]$ such that $z_i - z_0 s_0 s_{i^*} \neq 0$. Then the sum can be rewritten as

$$\langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_{i^*}\rangle \left|\langle \boldsymbol{w}_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle\right| + \sum_{i \neq i^*} \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i\rangle \left|\langle \boldsymbol{w}_i, (z_i - z_0 s_0 s_i)\boldsymbol{x}\rangle\right|. \tag{6}$$

We now show that the first term puts small probability mass over any short interval independent of the value of the summation over $i \neq i^*$. This gives a lower bound on the absolute value of the last coordinate by considering an interval around the negative of the second term.

Since $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$ implies $(z_0 s_0 s_{i^*}, \boldsymbol{z}) \in \mathcal{Z}_R$ for either of the two values of s_0 and s_{i^*}, the third condition in $\mathcal{E}_\delta$ gives

$$\left|\langle \boldsymbol{w}_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle\right| \;\geq\; \|\boldsymbol{x}\|_2^2 \cdot \frac{\pi}{2} \cdot \left(\frac{\delta}{2d|\mathcal{Z}_R|}\right)^2 .$$

Since $\boldsymbol{W}^{-1}$ is full rank, there is a component of $\tilde{\boldsymbol{w}}_{i^*}$ which is orthogonal to the span of $\{\tilde{\boldsymbol{w}}_i\}_{i \neq i^*}$. We write this as

$$\boldsymbol{u} \;=\; \tilde{\boldsymbol{w}}_{i^*} + \sum_{i \neq i^*} a_i \boldsymbol{w}_i$$

where $\langle \boldsymbol{u}, \boldsymbol{w}_i \rangle = 0$ for all $i \neq i^*$. Now let a_i for $i \neq i^*$ be the coefficients above, and let $a_{i^*} := 1$. Then $\boldsymbol{u} = \boldsymbol{W}^{-1}\boldsymbol{a}$ for $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)$, and thus

$$\|\boldsymbol{u}\|_2 \;\geq\; \frac{1}{2\sqrt{d} + \sqrt{2\ln(4/\delta)}} \cdot \|\boldsymbol{a}\|_2 \;\geq\; \frac{1}{2\sqrt{d} + \sqrt{2\ln(4/\delta)}} ,$$

where the first inequality follows from

$$\sigma_d(W^{-1}) \;\geq\; \frac{1}{2\sqrt{d} + \sqrt{2\ln(4/\delta)}}$$

on the event $\mathcal{E}_\delta$ and the second inequality from $a_{i^*} = 1$.

Thus Eq. (6) can be rewritten as the sum of two independent terms

$$\langle \boldsymbol{w}_{d+1}, \boldsymbol{u}\rangle \left|\langle \boldsymbol{w}_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle\right| + \sum_{i \neq i^*} \left(\langle \boldsymbol{w}_{d+1}, -a_i \tilde{\boldsymbol{w}}_i\rangle |\langle \boldsymbol{w}_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle + \langle \boldsymbol{w}_{d+1}, \tilde{\boldsymbol{w}}_i\rangle \left|\langle \boldsymbol{w}_i, (z_i - z_0 s_0 s_i)\boldsymbol{x}\rangle\right|\right) \quad (7)$$

The first term, $\langle \boldsymbol{w}_{d+1}, \boldsymbol{u}\rangle \left|\langle \boldsymbol{w}_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle\right|$, has distribution $\mathrm{N}(0, \sigma^2)$, where

$$\begin{aligned} \sigma^2 \;&\geq\; \|\boldsymbol{u}\|_2 |\langle w_{i^*}, (z_{i^*} - z_0 s_0 s_{i^*})\boldsymbol{x}\rangle| \\ &\geq\; \frac{\|\boldsymbol{x}\|_2^2 \frac{\pi}{2} \left(\frac{\delta}{2d|\mathcal{Z}_R|}\right)^2}{\left(2\sqrt{d} + \sqrt{2\ln(4/\delta)}\right)^2} . \end{aligned}$$

The event that Eq. (6) is small is when the Gaussian distribution returns a value in the interval of length $2r_\delta$ centered around the second term. The probability of this event is no more than

$$\frac{1}{\sqrt{2\pi\sigma^2}} \cdot 2r_\delta \;\leq\; \frac{2r_\delta}{\pi} \cdot \frac{2\sqrt{d} + \sqrt{2\ln(4/\delta)}}{\|\boldsymbol{x}\|_2 \left(\frac{\delta}{2d|\mathcal{Z}_R|}\right)} \;\leq\; \delta$$

by the choice of r_δ. Therefore, with probability at least $1 - \delta$, the quantity in 6 is at least r_δ, so the contribution of the last coordinate to the norm of the lattice vector is at least $\beta^2 r_\delta^2$, so the norm

of this lattice vector is at least $\sqrt{z_0^2 + \|\boldsymbol{z}\|_2^2 + \beta^2 r_\delta^2}$. To complete the proof we note that for all $(z_0, \boldsymbol{z}) \notin \mathcal{Z}_R$, by definition the norm of $\|\boldsymbol{z}\|_2$ is at least R. ■

We now prove Theorem 12. By the choices of R and β and Lemma 15, every incorrect coefficient vector has norm at least $2^{d/2}\sqrt{d+1}$, so it will not be returned by the LLL algorithm. By Lemma 14 there exists a short vector with coefficients $(s_0, \boldsymbol{s})$, so LLL recovers the correct signs.

Appendix D. Omitted proofs

D.1. Proof of Proposition 1

Claim 1 *$\lambda(\boldsymbol{X}) \geq \min_{i\in[k]} \|\boldsymbol{x}_i - \boldsymbol{\Pi}_{(-i)}\boldsymbol{x}_i\|_2$ where $\boldsymbol{\Pi}_{(-i)}$ is the orthogonal projection to the span of $\{\boldsymbol{x}_j\}_{j\neq i}$.*

Proof Let $\boldsymbol{v}$ be a non-zero vector in the lattice with basis $\boldsymbol{X}$. Write $\boldsymbol{v} = \sum_{j=1}^k z_j \boldsymbol{x}_j$, where $z_1, z_2, \ldots, z_k \in \mathbb{Z}$. Pick any $i \in [k]$ such that $z_i \neq 0$, and let $\boldsymbol{r} := -\sum_{j\neq i} z_j \boldsymbol{x}_j$, so

$$\|\boldsymbol{v}\|_2^2 = \|z_i\boldsymbol{x}_i - \boldsymbol{r}\|_2^2 \geq \left\|z_i\boldsymbol{x}_i - \boldsymbol{\Pi}_{(-i)}z_i\boldsymbol{x}_i\right\|_2^2 = |z_i|\left\|\boldsymbol{x}_i - \boldsymbol{\Pi}_{(-i)}\boldsymbol{x}_i\right\|_2^2 \geq \left\|\boldsymbol{x}_i - \boldsymbol{\Pi}_{(-i)}\boldsymbol{x}_i\right\|_2^2 .$$

Above, the first inequality follows from the Pythagorean theorem, and the second inequality follows because $z_i \in \mathbb{Z} \setminus \{0\}$. ■

By Claim 1, it suffices to lower-bound the distance between $\boldsymbol{x}_i$ and the subspace spanned by $\{\boldsymbol{x}_j\}_{j\neq i}$, for every $i \in [k]$. So fix $i \in [k]$, and any non-zero vector $\boldsymbol{r}_i$ in the span of $\{\boldsymbol{x}_j\}_{j\neq i}$. Let the singular value decomposition of $\boldsymbol{X}$ be given by $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{d\times k}$ has orthonormal columns, $\boldsymbol{S} = \operatorname{diag}(\sigma_1(\boldsymbol{X}), \sigma_2(\boldsymbol{X}), \ldots, \sigma_k(\boldsymbol{X})) \succ \boldsymbol{0}$ is diagonal, and $\boldsymbol{V} \in \mathbb{R}^{k\times k}$ is orthogonal. Let $\boldsymbol{\alpha}_j \in \mathbb{R}^k$ denote the j-th column of $\boldsymbol{V}^\top$. Then $\boldsymbol{x}_i = \boldsymbol{U}\boldsymbol{S}\boldsymbol{\alpha}_i$, and there exists non-zero $\boldsymbol{\beta}_i \in \mathbb{R}^k$ orthogonal to $\boldsymbol{\alpha}_i$ such that $\boldsymbol{r}_i = \boldsymbol{U}\boldsymbol{S}\boldsymbol{\beta}_i$. Moreover,

$$\frac{\langle \boldsymbol{x}_i, \boldsymbol{r}_i\rangle^2}{\langle \boldsymbol{x}_i, \boldsymbol{x}_i\rangle\langle \boldsymbol{r}_i, \boldsymbol{r}_i\rangle} = \frac{\left(\boldsymbol{\alpha}_i^\top \boldsymbol{S}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{S}\boldsymbol{\beta}_i\right)^2}{\left(\boldsymbol{\alpha}_i^\top \boldsymbol{S}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{S}\boldsymbol{\alpha}_i\right)\left(\boldsymbol{\beta}_i^\top \boldsymbol{S}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{S}\boldsymbol{\beta}_i\right)} = \frac{\left(\boldsymbol{\alpha}_i^\top \boldsymbol{S}^2\boldsymbol{\beta}_i\right)^2}{\left(\boldsymbol{\alpha}_i^\top \boldsymbol{S}^2\boldsymbol{\alpha}_i\right)\left(\boldsymbol{\beta}_i^\top \boldsymbol{S}^2\boldsymbol{\beta}_i\right)} .$$

By Wielandt's inequality (Horn and Johnson, 1985, 7.4.34), the ratio is bounded above by

$$\left(\frac{\sigma_1(\boldsymbol{S}^2)/\sigma_k(\boldsymbol{S}^2) - 1}{\sigma_1(\boldsymbol{S}^2)/\sigma_k(\boldsymbol{S}^2) + 1}\right)^2 = \left(\frac{\kappa(\boldsymbol{X})^2 - 1}{\kappa(\boldsymbol{X})^2 + 1}\right)^2 =: \phi .$$

By the Pythagorean theorem, the distance between $\boldsymbol{x}_i$ and the span of $\boldsymbol{r}_i$ is

$$\|\boldsymbol{x}_i\|_2 \left(1 - \frac{\langle \boldsymbol{x}_i, \boldsymbol{r}_i\rangle^2}{\langle \boldsymbol{x}_i, \boldsymbol{x}_i\rangle\langle \boldsymbol{r}_i, \boldsymbol{r}_i\rangle}\right)^{1/2} \geq \|\boldsymbol{x}_i\|_2 \sqrt{1-\phi} .$$

Since this holds for any $\boldsymbol{r}_i$ in the span of $\{\boldsymbol{x}_j\}_{j\neq i}$, the distance between $\boldsymbol{x}_i$ and the span of $\{\boldsymbol{x}_j\}_{j\neq i}$ is also at least

$$\|\boldsymbol{x}_i\|_2 \sqrt{1-\phi} = \|\boldsymbol{x}_i\|_2 \cdot \frac{2\kappa(\boldsymbol{X})}{\kappa(\boldsymbol{X})^2 + 1} .$$

The claim in Proposition 1 follows.

D.2. Proof of Lemma 3

It suffices to show the following probability bounds: (i) $\Pr(\sigma_d(\boldsymbol{W}) \leq \delta/(4\sqrt{d})) \leq \delta/4$; (ii) $\Pr(\|\boldsymbol{W}\|_2 > 2\sqrt{d} + \sqrt{2\ln(4/\delta)}) \leq \delta/4$; (iii) $\Pr(\text{Eq. (5) does not hold}) \leq \delta/(2dk|\mathcal{Z}_R|)$ for each $i \in [d]$, $j \in [k]$, and $(z_0, \boldsymbol{z}) \in \mathcal{Z}_R$ such that $|z_{i,j} - z_0| + \sum_{j' \neq j} |z_{i,j'}| > 0$. Combining these bounds with a union bound proves the claim.

The first two bounds follow from Theorem 9. The third requires Proposition 10 and the observation that the inner product in Eq. (5) is distributed as $\mathrm{N}(0, \|\boldsymbol{v}\|_2^2)$, where $\boldsymbol{v} := (z_{i,j} - z_0)\boldsymbol{x}_{\pi_i(j)} + \sum_{j' \neq j} z_{i,j'}\boldsymbol{x}_{\pi_i(j')}$. The condition on $(z_0, \boldsymbol{z})$ implies that $\boldsymbol{v}$ is a non-zero vector in the lattice $\Lambda(\boldsymbol{X})$, which has $\|\boldsymbol{v}\|_2 \geq \lambda(\boldsymbol{X})$ by definition.

D.3. Proof of Proposition 6

Let $\tilde{\boldsymbol{M}} := [\boldsymbol{a}|\boldsymbol{b}] \in \mathbb{R}^{d\times 2}$. The non-zero singular values of the matrix $\boldsymbol{M} = [\boldsymbol{a}+\boldsymbol{b}|\boldsymbol{a}-\boldsymbol{b}]$ are the same as the square-roots of the non-zero eigenvalues of

$$\boldsymbol{M}\boldsymbol{M}^\top \;=\; 2\boldsymbol{a}\boldsymbol{a}^\top + 2\boldsymbol{b}\boldsymbol{b}^\top \;=\; 2\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}^\top .$$

This matrix, in turn, has the same non-zero eigenvalues as the matrix

$$2\tilde{\boldsymbol{M}}^\top\tilde{\boldsymbol{M}} \;=\; 2\begin{bmatrix} \|\boldsymbol{a}\|_2^2 & \langle \boldsymbol{a}, \boldsymbol{b}\rangle \\ \langle \boldsymbol{b}, \boldsymbol{a}\rangle & \|\boldsymbol{b}\|_2^2 \end{bmatrix} .$$

The eigenvalues $\lambda_1 \geq \lambda_2$ of this matrix can be computed explicitly:

$$\begin{aligned}
\lambda_1 \;&=\; \|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2 + \sqrt{\left(\|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2\right)^2 - 4\left(\|\boldsymbol{a}\|_2^2\|\boldsymbol{b}\|_2^2 - \langle \boldsymbol{a}, \boldsymbol{b}\rangle^2\right)} , \\
\lambda_2 \;&=\; \|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2 - \sqrt{\left(\|\boldsymbol{a}\|_2^2 + \|\boldsymbol{b}\|_2^2\right)^2 - 4\left(\|\boldsymbol{a}\|_2^2\|\boldsymbol{b}\|_2^2 - \langle \boldsymbol{a}, \boldsymbol{b}\rangle^2\right)} .
\end{aligned}$$

Their ratio is

$$\frac{\lambda_1}{\lambda_2} \;=\; \frac{1 + \sqrt{1 - \frac{4\sin^2(\theta)}{\left(r+1/r\right)^2}}}{1 - \sqrt{1 - \frac{4\sin^2(\theta)}{\left(r+1/r\right)^2}}} ,$$

where $r = \|\boldsymbol{a}\|_2 / \|\boldsymbol{b}\|_2$, and θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$. The quantity $4\sin^2(\theta)/\left(r + 1/r\right)^2$ is always in the interval $[0, 1]$. A Taylor series expansion argument shows that

$$\frac{1 + \sqrt{1-x}}{1 - \sqrt{1-x}} \;\leq\; \frac{4}{x} , \quad x \in [0, 1] ,$$

so we conclude

$$\frac{\sigma_1(\boldsymbol{M})}{\sigma_2(\boldsymbol{M})} \;=\; \sqrt{\frac{\lambda_1}{\lambda_2}} \;\leq\; \frac{r + 1/r}{\left|\sin(\theta)\right|} .$$

D.4. Proof of Proposition 7

By homogeneity, we may assume $\|\boldsymbol{u}\|_2 = 1$. Let $g := \langle \boldsymbol{w}, \boldsymbol{u} \rangle \sim \mathrm{N}(0,1)$, and let $\boldsymbol{y} := \boldsymbol{w} - g\boldsymbol{u}$. Observe that g and $\boldsymbol{y}$ are independent, and

$$\mathbb{E}\boldsymbol{y} \;=\; \boldsymbol{0}\,, \qquad \mathbb{E}\boldsymbol{y}^{\otimes 2} \;=\; \boldsymbol{I}_d - \boldsymbol{u}^{\otimes 2} \;=\; \sum_{j=1}^{d} \boldsymbol{e}_j^{\otimes 2} - \boldsymbol{u}^{\otimes 2}\,, \qquad \mathbb{E}\boldsymbol{y}^{\otimes 3} \;=\; \boldsymbol{0}\,.$$

Using these facts, we have

$$\mathbb{E}\langle \boldsymbol{w}, \boldsymbol{u} \rangle^2 \left(\boldsymbol{w}^{\otimes 2} - \boldsymbol{I}_d \right) \;=\; \mathbb{E}g^2 \left(g^2 \boldsymbol{u}^{\otimes 2} + \boldsymbol{y}^{\otimes 2} - \boldsymbol{I}_d \right) \;=\; \mathbb{E}g^4 \boldsymbol{u}^{\otimes 2} - \mathbb{E}g^2 \boldsymbol{u}^{\otimes 2} \;=\; 2\boldsymbol{u}^{\otimes 2}\,,$$

$$\begin{aligned}
\mathbb{E}\langle \boldsymbol{w}, \boldsymbol{u} \rangle^3 \boldsymbol{w}^{\otimes 3} \;&=\; \mathbb{E}g^3 \left(g\boldsymbol{u} + \boldsymbol{y} \right)^{\otimes 3} \\
&=\; \mathbb{E}g^3 \left(g^3 \boldsymbol{u}^{\otimes 3} + g \left(\boldsymbol{u} \otimes \boldsymbol{y} \otimes \boldsymbol{y} + \boldsymbol{y} \otimes \boldsymbol{u} \otimes \boldsymbol{y} + \boldsymbol{y} \otimes \boldsymbol{y} \otimes \boldsymbol{u} \right) \right) \\
&=\; \mathbb{E}g^6 \boldsymbol{u}^{\otimes 3} + \mathbb{E}g^4 \sum_{j=1}^{d} \left(\boldsymbol{u} \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_j + \boldsymbol{e}_j \otimes \boldsymbol{u} \otimes \boldsymbol{e}_j + \boldsymbol{e}_j \otimes \boldsymbol{e}_j \otimes \boldsymbol{u} - 3\boldsymbol{u}^{\otimes 3} \right) \\
&=\; 6\boldsymbol{u}^{\otimes 3} + 3\mathcal{T}(\boldsymbol{u})\,, \\
\mathbb{E}\langle \boldsymbol{w}, \boldsymbol{u} \rangle^3 \boldsymbol{w} \;&=\; \mathbb{E}g^3 \left(g\boldsymbol{u} + \boldsymbol{y} \right) \;=\; 3\boldsymbol{u}\,,
\end{aligned}$$

so $\mathbb{E}\langle \boldsymbol{w}, \boldsymbol{u} \rangle^3 \left(\boldsymbol{w}^{\otimes 3} - \mathcal{T}(\boldsymbol{w}) \right) = 6\boldsymbol{u}^{\otimes 3}$. This proves the claims in Proposition 7.