Localization in 1D non-parametric latent space models from pairwise affinities

Christophe Giraud, Yann Issartel, Nicolas Verzelen

To cite this version:

Christophe Giraud, Yann Issartel, Nicolas Verzelen. Localization in 1D non-parametric latent space models from pairwise affinities. Electronic Journal of Statistics , 2023, 17 (1), pp.1587-1662. 10.1214/23-ejs2134 . hal-04166788

HAL Id: hal-04166788
https://hal.inrae.fr/hal-04166788

Submitted on 20 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Electronic Journal of Statistics
Vol. 17 (2023) 1587–1662
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2134

Localization in 1D non-parametric latent space models from pairwise affinities

Christophe Giraud

Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, CNRS, France.
e-mail: christophe.giraud@universite-paris-saclay.fr

Yann Issartel

CREST - ENSAE,
Télécom Paris, Institut Polytechnique de Paris, France.
e-mail: yann.issartel@telecom-paris.fr

Nicolas Verzelen

INRAE, Montpellier SupAgro, MISTEA, Univ. Montpellier, France.
e-mail: nicolas.verzelen@inrae.fr

Abstract: We consider the problem of estimating latent positions in a one-dimensional torus from pairwise affinities. The observed affinity between a pair of items is modeled as a noisy observation of a function $f(x_i^*, x_j^*)$ of the latent positions x_i^*, x_j^* of the two items on the torus. The affinity function f is unknown, and it is only assumed to fulfill some shape constraints ensuring that $f(x, y)$ is large when the distance between x and y is small, and vice-versa. This non-parametric modeling offers a good flexibility to fit data. We introduce an estimation procedure that provably localizes all the latent positions with a maximum error of the order of $\sqrt{\log(n)/n}$, with high-probability. This rate is proven to be minimax optimal. A computationally efficient variant of the procedure is also analyzed under some more restrictive assumptions. Our general results can be instantiated to the problem of statistical seriation, leading to new bounds for the maximum error in the ordering.

Received February 2023.

1. Introduction

1.1. 1D latent localization problem

We consider the 1D latent localization problem, where we seek to recover the 1D latent positions of n objects from pairwise similarity measurements. Such problems arise in archeology for relative dating of objects or graves [34], in 2D-tomography for angular synchronization [10, 37], in bioinformatics for reads alignment in *de novo* sequencing [32], in computer science for time synchronization in distributed networks [14, 22], or in matchmaking problems [6]. The data are collected as a $n \times n$ symmetric matrix $[A_{ij}]_{1 \leq i,j \leq n}$, called affinity matrix, which provides similarity measurements between pairs of objects. These similarity measurements A_{ij} can be real valued scores, or they can be binary pieces of information, as when the matrix A encodes a network structure.

In 1D latent space models [24], the affinity matrix is assumed to be sampled as follows. The distribution is parametrized by a 1D metric space $(\mathcal{X}, d)$, some (possibly random) latent positions $(x_1^*, \dots, x_n^*) \in \mathcal{X}^n$ and an affinity function $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Then, conditionally on $(x_1^*, \dots, x_n^*)$, the upper-diagonal entries A_{ij} of the affinity matrix are sampled independently, with conditional mean $f(x_i^*, x_j^*)$. The affinity $f(x_i^*, x_j^*)$ is typically assumed to decrease as the metric distance $d(x_i^*, x_j^*)$ increases. In particular, close points x_i^* and x_j^* share a high affinity, whereas distant points share a small affinity. These latent space models encompass many classical models, as exemplified in the next paragraphs.

Example 1: Random Geometric Graph [21, 30, 13, 11]. We observe a random graph with n nodes labelled by $\{1, \dots, n\}$. The graph is encoded into an adjacency matrix $A \in \{0,1\}^{n \times n}$, by setting $A_{ij} = 1$ if there is an edge between nodes i and j, and $A_{ij} = 0$ otherwise. Let $\mathcal{C}$ denote the unit sphere in $\mathbb{R}^2$ endowed with the geodesic distance d. In the circular random geometric graph model, the edges are sampled are sampled independently, with probability $\mathbb{P}[A_{ij} = 1] = g(d(x_i^*, x_j^*))$, where $g : [0, \pi] \mapsto [0,1]$ is a non-increasing function and $x_1^*, \dots, x_n^* \in \mathcal{C}$ are the latent positions of the nodes on the sphere. This random graph model is therefore an instance of 1D latent space model where $A_{ij} \in \{0,1\}$, $\mathcal{X} = \mathcal{C}$ and $f(x_i^*, x_j^*) = g(d(x_i^*, x_j^*))$.

Example 2: Graphons and f-Random Graphs [12, 27]. The class of f-random graph models, also called graphon models, encompasses all the distributions on random graphs that are invariant by permutation of nodes. It is parametrized by the set of measurable functions $f : [0,1] \times [0,1] \to [0,1]$. The adjacency matrix A of the graph is sampled as follows. First, n latent positions $x_1^*, \dots, x_n^*$ are sampled i.i.d. uniformly on $[0,1]$. Then, conditionally on $x_1^*, \dots, x_n^*$, the edges are sampled independently, with conditional probability $\mathbb{P}[A_{ij} = 1 | x_1^*, \dots, x_n^*] = f(x_i^*, x_j^*)$. The f-random graph model is then an instance of 1D latent space model where $A_{ij} \in \{0,1\}$, and $\mathcal{X} = [0,1]$. Unless some additional constraints are imposed on the shape of f, the affinity $f(x_i^*, x_j^*)$ may vary arbitrarily with the distance $|x_i^* - x_j^*|$.

Example 3: R-Matrices and Statistical Seriation. A Robinson matrix (R-matrix) is any symmetric matrix $B \in \mathbb{R}^{n \times n}$ whose entries decrease when moving away from the diagonal, i.e. such that $B_{i,j} \geq B_{i+1,j}$ and $B_{i,j} \geq B_{i,j-1}$, for all $1 \leq j \leq i \leq n$. A matrix F is called a pre-R matrix, when there exists a permutation $\sigma \in \Sigma_n$ of $\{1, \dots, n\}$, such that $F_\sigma = [F_{\sigma(i),\sigma(j)}]_{i,j}$ is an R-matrix. The noisy seriation problem [15] amounts to find, from a noisy observation of a pre-R matrix F, a permutation σ^* such that F_{σ^*} is a R-matrix. This problem appears in genomic sequencing [20], in interval graph identification [16], and in envelope reduction for sparse matrices [5]. This problem can be recast in the latent space terminology using $\mathcal{X} = \{1, \dots, n\}$, $x_i^* = \sigma^*(i)$ and the affinity function $f(x_i^*, x_j^*) = F_{\sigma^*(i),\sigma^*(j)}$. Since F_{σ^*} is a R-matrix, the function $f(x_i^*, x_j^*)$ is decreasing with the distance $|x_i^* - x_j^*|$.

Example 4: Toroidal R-Matrices and Toroidal Seriation. Consider the set $\{1, \dots, n\}$ as a torus with the corresponding distance $d(i, j) = \min(|j-i|, |n+i-j|)$ for any $1 \leq i, j \leq n$. A toroidal R-matrix is any symmetric matrix B whose entries decrease when moving away from the diagonal with respect to the toroidal distance: $B_{i,j} \geq B_{i+1,j}$ when $d(i, j) < d(i+1, j)$ and $B_{i,j} \geq B_{i,j+1}$ when $d(i, j) < d(i, j+1)$. As in Example 3 above, a pre-toroidal R-matrix is defined as a permutation of a toroidal R-matrix and the statistical seriation model is defined analogously [33]. Again, we can recast this model as a latent space model on $\mathcal{X} = \{1, \dots, n\}$ endowed with the toroidal distance. Alternatively, we can also rewrite it as a latent space model on the regular grid $\mathcal{C}_n$ of the unit sphere $\mathcal{C}$ corresponding to the n-th unit roots, endowed with the geodesic distance on $\mathcal{C}$.

In the following, we assume that we observe a symmetric matrix $[A_{ij}]_{1 \leq i,j \leq n}$ of pairwise affinity measurements, with $A_{ii} = 0$ (by convention) and

$$A_{ij} = f(x_i^*, x_j^*) + E_{ij}, \quad \text{for } 1 \leq i < j \leq n,$$

where

(i) $x_1^*, \dots, x_n^*$ are n unobserved latent positions spread on the unit sphere $\mathcal{C}$ in $\mathbb{R}^2$,
(ii) $f : \mathcal{C} \times \mathcal{C} \to [0, 1]$ is unobserved, symmetric, decreasing with the geodesic distance $d(x, y)$, and
(iii) $[E_{ij}]_{1 \leq i < j \leq n}$ are some independent sub-Gaussian random variables.

This non-parametric framework is very flexible for fitting pairwise affinity data. It encompasses the circular random geometric graph model (Example 1) and the toroidal statistical seriation model (Example 4).

Our overall goal is to recover the n-tuple of latent positions $\mathbf{x}^* = (x_1^*, \dots, x_n^*) \in \mathcal{C}^n$, with some high-confidence, simultaneously for all individual positions x_i^*. As the global error of an estimator $\hat{\mathbf{x}}$, say $d_2(\hat{\mathbf{x}}, \mathbf{x}^*) = \sqrt{\sum_{i=1}^n d(\hat{x}_i, x_i^*)^2}$, provides limited information on each individual error $d(\hat{x}_i, x_i^*)$, we focus instead on the maximum error

$$d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*) = \max_{i=1,\dots n} d(\hat{x}_i, x_i^*). \tag{1}$$

We propose some estimators $\hat{\mathbf{x}}$ achieving, with high-probability, a maximum error $d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*)$ of the order of $\sqrt{\log(n)/n}$, under the assumptions that the latent positions $x_1^*, \dots, x_n^*$ are sufficiently spread on $\mathcal{C}$ and some shape conditions relative to the decreasing of $f(x, y)$ with $d(x, y)$. The $\sqrt{\log(n)/n}$-rate of estimation is shown to be optimal. To the best of our knowledge, these are the first optimal results on maximum error $d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*)$ in latent space models with unknown and non-parametric affinity function f.

1.2. Our contribution

As explained above, our overall goal is to recover the n-tuple of latent positions $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ with a control on the maximum error (1). Unfortunately, this program cannot be carried out literally, as the latent positions are not identifiable from the distribution of the data. Indeed, for any bijective map $\varphi : \mathcal{C} \to \mathcal{C}$, we have $f(x, y) = f \circ \varphi^{-1}(\varphi(x), \varphi(y))$ for all $x, y \in \mathcal{C}$, with the notation $f \circ \varphi^{-1}(x, y) := f(\varphi^{-1}(x), \varphi^{-1}(y))$. Even if we would enforce some strong shape constraints, like $f(x, y) = 1 - \alpha d(x, y)$ with $\alpha > 0$, since $f(x, y) = f(Qx, Qy)$ for any orthogonal transformation Q of $\mathcal{C}$, the distribution of the data would still be invariant by orthogonal transformation of the latent positions. Hence, we face a delicate identifiability issue. This identifiability issue is fully explained and tackled in Section 2.2. Informally, our remedy is to provide some estimators $\hat{\mathbf{x}}$ which are, under some assumptions, at the distance $d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*) = O(\sqrt{\log(n)/n})$ of some specific representative $\mathbf{x}^*$ of the latent positions.

Our shape assumption (ii) on the affinity function f ensures that the matrix $[f(x_i, x_j)]_{i,j=1,\ldots,n}$ is (approximately) a toroidal pre-R matrix. We observe that the constant function $f(x, y) = 1$ fulfills assumption (ii), and that for this specific function there is no hope to recover any information on the latent positions, even in the noiseless case. To circumvent this issue, we introduce a *bi-Lipschitz* assumption, detailed in Section 2.1, constraining the decay of f with d. In the specific case of the random geometric graph model with g continuously differentiable, this condition merely amounts to require $g'(x)$ to be bounded away from 0.

Our estimation procedures proceed in two main stages:
(1) we start with an initial localization with a global control in $d_1(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^n d(x_i, y_i)$ distance,
(2) then, for each point, we refine this first estimator to get a control in d_∞ distance.

In order to avoid some nasty statistical dependencies between the two stages, we use a sample splitting scheme ensuring that, at the second stage, the refinement uses data independent from those used at the first stage.

Let $S \subset \{1, \ldots, n\}$ be a subset of indices sampled uniformly at random, and $\overline{S} = \{1, \ldots, n\} \setminus S$. At the second step, the refined estimator $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ of $\mathbf{x}_{\overline{S}}^* := (x_i^*)_{i \in \overline{S}}$, can take as input any initial estimator $\hat{\mathbf{x}}_S^{(1)}$ of $\mathbf{x}_S^* = (x_i^*)_{i \in S}$. This second step has a polynomial computational complexity and, under appropriate assumptions, it fulfills with high-probability

$$d_\infty(\hat{\mathbf{x}}_{\overline{S}}^{(2)}, \mathbf{x}_{\overline{S}}^*) \leq C \left(\frac{d_1(\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*)}{n} \right) \vee \sqrt{\frac{\log(n)}{n}},$$

for some specific representative $\mathbf{x}^*$ of the latent positions. Hence, in order to get the desired bound $d_\infty(\hat{\mathbf{x}}_{\overline{S}}^{(2)}, \mathbf{x}_{\overline{S}}^*) = O\left(\sqrt{\log(n)/n}\right)$, we need an initial control $d_1(\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*) = O\left(\sqrt{n \log(n)}\right)$. We propose two estimators fulfilling this requirement:

(a) a first one, which requires no additional assumptions, but which has a super-polynomial computational complexity;
(b) a second one, adapted from [33], which has a polynomial computational complexity, but for which we prove a $O\big(\sqrt{n\log(n)}\big)$ control only for a class of random geometric graphs.

Repeating the sampling of S and merging the resulting estimators, we then get an estimator $\hat{\mathbf{x}}$ achieving, with high-probability and under appropriate assumptions, $d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*) = O(\sqrt{\log(n)/n})$ for a specific representative $\mathbf{x}^*$ of the latent positions. A matching lower bound is also derived, proving the optimality of the $\sqrt{\log(n)/n}$ rate. The significance of the improvement offered by the refinement step, and the impact of the sample splitting on the localization error are investigated numerically.

1.3. Related work

In the last decade, the analysis of interaction data has given rise to numerous works in machine learning and statistics. Most of these works handle cases where the affinity function f is either known or belong to a known parametric model. There is a long standing debate on the validity of such a rigid modeling [4]. Our modeling assumptions, with only shape constraints on f, offers a more flexible setting to fit data.

Latent points estimation in random geometric graphs Random geometric graphs have attracted a lot of attraction as a simple model for wireless communications or internet [21, 30]. In the most classical setting, $A_{ij} = f(x_i^*, x_j^*) = \mathbf{1}_{\|x_i^* - x_j^*\| \le r}$ for some $r > 0$. The problem of estimating the latent positions $x_1^*, \dots, x_n^*$ in a square of $\mathbb{R}^2$ has been tackled by [13]. Compared to us, they consider the noiseless setting, where $E_{ij} = 0$, with the affinity map f belongs to 1-dimensional parametric model. The problem of latent positions localization has also been investigated in the random dot-product graph [38, 28, 2], where, conditionally to the latent positions, the entries A_{ij} of the adjacency matrix are independent Bernoulli random variables with mean $f(x_i^*, x_j^*) = \langle x_i^*, x_j^* \rangle$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{R}^d$. In this case, the function f is known and the results are of an asymptotic nature.

Phase synchronization problems The phase synchronization problem [37] amounts to estimating unknown angles $\theta_1, \dots, \theta_n$ from noisy measurements of $\theta_i - \theta_j \mod 2\pi$. A version of this problem is when we seek to retrieve $x_1^*, \dots, x_n^*$, which are spread on the unit complex sphere $\mathbb{C}_1 = \{x \in \mathbb{C} : |x| = 1\}$, from noisy observations of $x_i^* \overline{x}_j^* = e^{\iota(\theta_i - \theta_j)}$. In this model, some minimax ℓ^2-bounds on the localization error have been obtained by [19], without assumptions on the latent positions $x_1^*, \dots, x_n^*$. Such a model is close to the pairwise affinity model on $\mathcal{C}$. The main differences compared to our setting is that we focus on d_∞-bounds, with an unknown function f, which does not have the affinity shape (ii).

Skills estimation in the Bradley-Terry model In the Bradley-Terry model [6], the observations A_{ij} are independent Bernoulli outcomes with mean $f(x_i^*, x_j^*) = \sigma(x_i^* - x_j^*)$, where $x_i^* \in \mathbb{R}$ represents the skill of individual i and $\sigma(x) = e^x/(1+e^x)$ is the sigmoid function. The estimation of the skills by a spectral algorithm, or two-steps variants of it, has received a lot of attention recently [29, 9, 8]. In particular, building on the structure of the problem, rate-minimax ℓ_∞-bound have been derived for the spectral algorithm in the Bradley-Terry model when the skills belong to a compact set, possibly with missing at random observations.The Bradley-Terry model is a special instance of the 1D-latent space model. Compared to our setting, the function f is known and it does not fulfill the affinity properties (for example, it is not symmetric).

Seriation from pairwise affinity Given a pre-R matrix F, the seriation problem seeks to find the latent order σ^* such that F_{σ^*} is a R-matrix. For this noiseless version of Example 3, efficient algorithms have been proposed using convex optimization [15], or spectral methods [3]. The exact seriation problem has been solved on toroidal R-matrices in the noiseless case [33], by using a spectral algorithm. A perturbation analysis has also been sketched in [33]. As a byproduct of our analysis, we provide some more explicit recovery bounds in our specific setting with noisy observations. Closer to our contribution, Jannssen and Smith [25] observe a noisy version of a pre-R matrix and, under some assumptions on the affinity function f, learn a permutation that satisfies $\max_{i\in[n]} |\widehat{\sigma}_i - \sigma_i^*| \leq C\sqrt{n}\log^5(n)$. Although their assumptions on f are not directly comparable to ours, the localization rates are (up to logarithmic factors) comparable to ours. We refer to the discussion below Corollary 3.4 for more details.

Two-step methods for latent space models Our work is related to the global-to-local estimation strategy, that was originally introduced in Stochastic Block Models and more generally in clustering analysis [26, 17, 43], for the purpose of deriving sharp recovery bounds with polynomial time procedures. The general idea is to build upon an initial estimator that satisfies a certain (weak) consistency condition, and then apply greedy-type procedures (e.g. Lloyd's algorithm) to obtain minimax recovery bounds. This approach turned out to be fruitful in various latent space problems with discrete structure [7, 18] and our procedure can be interpreted as one instance of this strategy in a non-parametric setting with a continuous latent space.

1.4. Notation and organization of the paper

In the sequel, $C, C', C'' > 0$ denote numerical constants that may change from line to line. For two functions or sequences x and y, we write $x \lesssim y$ (resp. $x \gtrsim y$) when, for some numerical constant $C > 0$, we have $x \leq Cy$ (resp. $x \geq Cy$). The maximum (resp. minimum) of x and y is denoted by $x \vee y$ (resp. $x \wedge y$). For any $x > 0$, we write $\lfloor x \rfloor$ for its integer part, and $[x]$ for the set of integers $[x] = \{1, \dots, \lfloor x \rfloor\}$. For $q \geq 1$, the entry-wise l_q norm of a matrix $F = (f_{ij})$ is denoted by $\|F\|_q = (\sum_{ij} |f_{ij}|^q)^{1/q}$, the i^{th}-row of F is denoted by F_i, and the Frobenius scalar product between two matrices F and G is denoted by $\langle F, G \rangle$. Let Σ_n be the collection of permutations of $[n]$. For any permutation $\sigma \in \Sigma_n$, and for any n-tuple $\mathbf{x}$ of size n, we define $\mathbf{x}_\sigma$ as the permuted n-tuple $\mathbf{x}_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$.

Assimilating points in the unit sphere $\mathcal{C}$ of $\mathbb{R}^2$ to complex numbers with unit norm, we can represent $x \in \mathcal{C}$ by $x = e^{\iota \underline{x}}$, with $\underline{x} \in [0, 2\pi)$. We call henceforth argument of $x \in \mathcal{C}$ the real number $\underline{x}$. The geodesic distance $d(x, y)$ on $\mathcal{C}$ can be conveniently defined in terms of the arguments of x and y

$$d(x, y) = |\underline{x} - \underline{y}| \wedge \left(2\pi - |\underline{x} - \underline{y}|\right). \tag{2}$$

For any positive integer k, we define the regular grid $\mathcal{C}_k = \{1, e^{\iota 2\pi/k}, \dots, e^{\iota 2\pi(k-1)/k}\}$, which plays an important role in our analysis and algorithms. We denote by $\mathcal{O}$ the orthogonal group of $\mathbb{R}^2$ made of rotations and reflections, and for any n-tuple $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{C}^n$, and any $Q \in \mathcal{O}$, we define $Q\mathbf{x} := (Qx_1, \dots, Qx_n)$. For two subsets S, S' of $\{1, \dots, n\}$, a matrix $A \in \mathbb{R}^{n \times n}$ and a n-tuple $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{C}^n$, we define $A_{SS'} = [A_{ij}]_{i \in S, j \in S'}$ and $\mathbf{x}_S = [x_i]_{i \in S}$. More generally, we denote by $\underline{\mathbf{x}}_S \in \mathcal{C}^S$ a $|S|$-tuple indexed by S. The complement of a set S, is denoted by $\overline{S}$.

In Section 2, we describe the statistical setting and we discuss thoroughly the identifiability issues. The main embedding procedure, called Localize-and-Refine, is presented in Section 3. A spectral variant of this procedure is introduced in Section 4, with an application to geometric models. In Section 6, we investigate numerically the usefulness of the sample splitting, and the significance of the improvement offered by the refinement step. We summarize our findings and discuss an open problem in Section 7. All the proofs are postponed to the appendices.

2. Model assumptions and identifiability issues

2.1. Statistical setting

We observe a realization of a symmetric random matrix $A \in \mathbb{R}^{n \times n}$, whose values on the diagonal are $A_{ii} = 0$. We denote by $F_{ij} = \mathbb{E}[A_{ij}]$ the mean value of A_{ij} and by $E_{ij} = A_{ij} - F_{ij}$ the centered random fluctuation. We assume that A has been generated by a latent space model on $\mathcal{C}$: there exist a n-tuple $\mathbf{x}^* = (x_1^*, \dots, x_n^*) \in \mathcal{C}^n$ and a function $f : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ such that $F_{ij} = f(x_i^*, x_j^*)$, so

$$A_{ij} = F_{ij} + E_{ij} = f(x_i^*, x_j^*) + E_{ij}, \quad \text{for} \quad 1 \leq i < j \leq n. \tag{3}$$

Both the function f and the latent positions $x_1^*, \dots, x_n^*$ are unknown. We emphasize that the latent positions $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ are assumed to be fixed[1] and we denote by $\mathbb{P}_{(\mathbf{x}^*,f)}$ the distribution of A. Let us describe our assumptions on the spreading of the latent positions $x_1^*,\dots,x_n^*$, the shape of f, and the random fluctuations E_{ij}.

---

[1] if they were random, our results would apply conditionally on the sampling of $\mathbf{x}^*$.

Spreading of the latent positions We have in mind that the latent positions are well spread over the unit sphere. We do not strictly enforce this condition, but our error bounds depend on how far the latent positions are from a regular position on $\mathcal{C}$. More precisely, let us denote by $\mathbf{\Pi}_n$ the set of regular positions on the unit sphere $\mathbf{\Pi}_n = \left\{\mathbf{x} = (e^{\iota 2\pi\sigma(j)/n})_{1\leq j\leq n} : \sigma \in \Sigma_n\right\}$. Our results involve the d_∞-distance of the n-tuple of latent positions $\mathbf{x}^*$ to the set $\mathbf{\Pi}_n$ of regular positions

$$d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) := \min_{\mathbf{y}\in\mathbf{\Pi}_n} \; d_\infty(\mathbf{x}^*, \mathbf{y}) \ , \tag{4}$$

with $d_\infty(\mathbf{x}^*, \mathbf{y})$ defined in (1).

Bi-Lipschitz shape of f As explained in the introduction, we have in mind that $f(x,y)$ decreases with the distance $d(x,y)$. Since there is no hope to recover the latent positions $\mathbf{x}^*$ when the function f is flat, we impose a minimal decreasing of $f(x,y)$ with the distance $d(x,y)$. We also require some Lipschitz continuity of f for our analysis. These two conditions on f are enforced by the Bi-Lipschitz condition described below.

Definition 2.1. ***Bi-Lipschitz functions.*** *For any fixed constants $c_e \geq 0$ and $0 < c_l \leq c_L$, we define $\mathcal{BL}[c_l, c_L, c_e]$ as the set made of all functions $f : \mathcal{C}^2 \to [0,1]$ that are symmetric (i.e. $f(x,y) = f(y,x)$ for all $x, y \in \mathcal{C}$) and that satisfy the two following conditions for all $x, y, y' \in \mathcal{C}$,*

$$|f(x,y) - f(x,y')| \leq c_L d(y,y') + \varepsilon_n \ ; \tag{5}$$

$$f(x,y') - f(x,y) \geq c_l\big(d(x,y) - d(x,y')\big) - \varepsilon_n \quad \textit{if } d(x,y) \geq d(x,y') \ , \tag{6}$$

with $\varepsilon_n = c_e\sqrt{\log(n)/n}$.

When $c_e = 0$, Condition (5) enforces Lipschitz continuity and Condition (6) enforces a minimal decreasing of $f(x,y)$ with $d(x,y)$. In the geometric case $f = g \circ d$ with $g : [0 : \pi] \to [0,1]$ continuously differentiable, these conditions hold when $-c_L \leq g'(t) \leq -c_l$ for all $t \in [0,\pi]$. For $c_e > 0$, the term ε_n in (5–6) can be interpreted as a possible small relaxation of a strict bi-Lipschitz property. In the remaining of the paper, we will assume that $f \in \mathcal{BL}[c_l, c_L, c_e]$ for some $c_e \geq 0$ and $0 < c_l \leq c_L$.

SubGaussian errors We assume that the entries E_{ij} for $1 \leq i < j \leq n$ of the noise matrix are independent and follow a subGaussian(1) distribution. It means that, for any matrix $B \in \mathbb{R}^{n\times n}$ and $t \geq 0$, we have

$$\mathbb{P}\left[\sum_{1\leq i<j\leq n} B_{ij}E_{ij} > t\sqrt{\sum_{1\leq i<j\leq n} B_{ij}^2}\right] \leq e^{-t^2/2}. \tag{7}$$

Since centered random variables taking values in $[-1,1]$ have a subGaussian(1) distribution, this setting encompasses the case where $A \in \{0,1\}^{n\times n}$ is the adjacency matrix of a random graph, whose distribution belongs to a latent space model on $\mathcal{C}$.

To keep the notation and the presentation simple, we assume henceforth that the sample size n is a multiple of 4, and we denote by n_0 the integer $n_0 = n/4$.

2.2. Identifiability issues

Our overall goal is to estimate the latent positions $\mathbf{x}^*$. Yet, in general, these latent positions are not identifiable from the distribution of A. Indeed, for any bijective $\varphi : \mathcal{C} \to \mathcal{C}$, we have $[F_{ij}]_{i<j} = \left[(f\circ\varphi^{-1})(\varphi(x_i^*), \varphi(x_j^*)\right]_{i<j}$, with the notation $f\circ\varphi^{-1}(x,y) := f(\varphi^{-1}(x), \varphi^{-1}(y))$. Hence, it is not possible to recover $\mathbf{x}^*$ from F and, unless $\mathbf{x}^*$ is identifiable from the distribution of the noise E, the n-tuple of latent positions $\mathbf{x}^*$ is not identifiable. Worse, F can be represented by many different couple $(\mathbf{x}, f)$. Hence, we face a serious identifiability issue. However, with the premise that the latent positions are well spread on $\mathcal{C}$, we can give a sensible meaning to our estimation objective. We explain progressively the issues that we face, in order to clarify the problem.

As a warm-up, let us assume in this paragraph that f is known. Even in this favorable case, there might exist some bijective $\varphi : \mathcal{C} \to \mathcal{C}$ such that $f = f\circ\varphi$, and hence $f(x_i^*, x_j^*) = f(\varphi(x_i^*), \varphi(x_j^*))$. For example, when $f = g\circ d$, we have $f = f\circ Q$ for any orthogonal transformation $Q \in \mathcal{O}$. In this last case, unless $\mathbf{x}^*$ is identifiable from the distribution of the noise E, the best that we can hope is to consistently estimate $\mathbf{x}^*$ in terms of the quasi-distance

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*), \quad \text{where } Q\mathbf{x} := (Qx_1, \ldots, Qx_n).$$

Let us come back to our setting where f is unknown. We define $\mathcal{R}[F, c_l, c_L, c_e]$ (or simply $\mathcal{R}[F]$) as the set of representations of F by n-tuples in $\mathcal{C}^n$ and bi-Lipschitz functions

$$\mathcal{R}[F, c_l, c_L, c_e] := \left\{(\mathbf{x}, f) \in \mathcal{C}^n \times \mathcal{BL}[c_l, c_L, c_e] : f(x_i, x_j) = F_{ij} \text{ for all } i < j\right\}. \tag{8}$$

We observe that for any $f \in \mathcal{BL}[c_l, c_L, c_e]$ and any $Q \in \mathcal{O}$, we have $f\circ Q^{-1} \in \mathcal{BL}[c_l, c_L, c_e]$. Hence, if $(\mathbf{x}^*, f) \in \mathcal{R}[F, c_l, c_L, c_e]$, then $\left\{(Q\mathbf{x}^*, f\circ Q^{-1}) : Q\in\mathcal{O}\right\} \subset \mathcal{R}[F, c_l, c_L, c_e]$. Are all the elements in $\mathcal{R}[F, c_l, c_L, c_e]$ of the form $(Q\mathbf{x}^*, f\circ Q^{-1})$, as in the case discussed above?

Let us first focus on the case where $(\mathbf{x}, f), (\mathbf{x}', f') \in \mathcal{R}(F)$ with $\mathbf{x}, \mathbf{x}' \in \mathbf{\Pi}_n$ and $c_e = 0$. The Proposition 2.3 below ensures that there exists $Q \in \mathcal{O}$ such that $\mathbf{x}' = Q\mathbf{x}$. Hence, if regular representations $(\mathbf{x}, f)$ exist in $\mathcal{R}[F]$, all the other regular representations are given by $(Q\mathbf{x}^*, f\circ Q^{-1})$ with $Q \in \mathcal{O}$ letting $\mathbf{\Pi}_n$ invariant.

This property breaks down when we move away from regular latent positions in $\mathbf{\Pi}_n$. Indeed, the next proposition shows that we can have $\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}')$

large even, when $\mathbf{x} \in \mathbf{\Pi}_n$ and $\mathbf{x}'$ is well spread on $\mathcal{C}$. More precisely, this property is shown for

$$\mathbf{x}' \in \mathcal{S}_{ev} := \Big\{\mathbf{x} \in \mathcal{C}^n : \sup_{z\in\mathcal{C}} \min_{i\in[n]} d(x_i, z) \leq 3\pi/n\Big\}.$$

Such a n-tuple $\mathbf{x}'$ is well spread on $\mathcal{C}$, since any $z \in \mathcal{C}$ is at a distance at most $3\pi/n$ from one of the x'_i.

Proposition 2.2. *Assume that* $F_{ij} = f(x_i, x_j)$ *for* $1 \leq i < j \leq n$, *with* $f(x, y) = 1 - d(x, y)/(2\pi)$ *and* $x_k = e^{\iota k2\pi/n}$ *for all* $k \in [n]$. *Then, there exists another representation* $(\mathbf{x}', f') \in \mathcal{R}[F, (3\pi)^{-1}, \pi^{-1}, 0]$, *with* $\mathbf{x}' \in \mathcal{S}_{ev}$, *such that*

$$\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}') \geq \pi/8 \ .$$

The representation $(\mathbf{x}', f')$ can be obtained from $(\mathbf{x}, f)$ by slightly stretching and contracting some pieces of the sphere $\mathcal{C}$. The detailed proof of this proposition is postponed to Appendix C. This result shows that the set of latent positions $\mathbf{x}'$ in $\mathcal{R}[F, (3\pi)^{-1}, \pi^{-1}, 0]$ is much richer than the set $\{Q\mathbf{x} : Q \in \mathcal{O}\}$ of orthogonal transformations of $\mathbf{x}$, since it includes some latent positions $\mathbf{x}'$ at constant d_∞-distance from this set. Yet, the next proposition shows that for any $\mathbf{x}, \mathbf{x}'$ in $\mathcal{R}[F]$, the d_∞-distance of $\mathbf{x}'$ to $\{Q\mathbf{x} : Q \in \mathcal{O}\}$ is controlled in terms of the d_∞-distance of $\mathbf{x}$ and $\mathbf{x}'$ to the set of regular positions $\mathbf{\Pi}_n$.

Proposition 2.3. *Let* F *be a symmetric matrix given by* $F_{ij} = f(x^*_i, x^*_j)$ *for* $1 \leq i < j \leq n$, *with* $\mathbf{x}^* \in \mathcal{C}^n$ *and* $f \in \mathcal{BL}[c_l, c_L, c_e]$ *for some* $c_e \geq 0$ *and* $0 < c_l \leq c_L$. *Then, there exists a constant* $C_{lLe} > 0$, *depending only on* c_e, c_l *and* c_L, *and such that, for any* $(\mathbf{x}, f), (\mathbf{x}', f') \in \mathcal{R}[F, c_l, c_L, c_e]$, *we have*

$$\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}') \leq C_{lLe}\left(d_\infty(\mathbf{x}, \mathbf{\Pi}_n) + d_\infty(\mathbf{x}', \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right). \tag{9}$$

If, in addition, $c_e = 0$ *and* $\mathbf{x}, \mathbf{x}' \in \mathbf{\Pi}_n$, *then there exists* $Q \in \mathcal{O}$ *such that* $\mathbf{x} = Q\mathbf{x}'$.

The proof of (9) can be found in Appendix C.2, whereas the proof of the second statement can be found in Appendix A. This result shows that if $\mathbf{x}$ and $\mathbf{x}'$ are close to $\mathbf{\Pi}_n$, then $\mathbf{x}$ is close to an orthogonal transformation of $\mathbf{x}'$. Hence, when we restrict to representations $(\mathbf{x}, f)$ and $(\mathbf{x}', f')$, with latent positions $\mathbf{x}$ and $\mathbf{x}'$ close to $\mathbf{\Pi}_n$, the identifiability issue becomes smoother.

Summarizing our discussion above, we have shown that:

(1) for any $(\mathbf{x}, f) \in \mathcal{R}[F]$, we have $\big\{(Q\mathbf{x}, f \circ Q^{-1}) : Q \in \mathcal{O}\big\} \subset \mathcal{R}[F]$;

(2) for any $(\mathbf{x}, f), (\mathbf{x}', f') \in \mathcal{R}[F]$, the d_∞-distance of $\mathbf{x}'$ to $\{Q\mathbf{x} : Q \in \mathcal{O}\}$ is bounded in terms of the d_∞-distance of $\mathbf{x}$ and $\mathbf{x}'$ to $\mathbf{\Pi}_n$.

Accordingly, to contain the phenomenon described in Proposition 2.2, we will focus henceforth on the representations $(\mathbf{x}, f)$ which are the closest to $\mathbf{\Pi}_n$.

Problem formulation Let us explain our estimation strategy, in light of the above discussion. In order to circumvent the identifiability issues, we focus on the representations $(\mathbf{x}, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ whose latent positions are the closest to $\mathbf{\Pi}_n$, i.e. we focus on the following set[2] of representations

$$\mathcal{R}_{\mathbf{\Pi}_n}[F] = \mathcal{R}_{\mathbf{\Pi}_n}[F, c_l, c_L, c_e] := \underset{(\mathbf{x},f)\in\mathcal{R}[F,c_l,c_L,c_e]}{\operatorname{argmin}} d_\infty(\mathbf{x}, \mathbf{\Pi}_n). \tag{10}$$

Our goal is then to build an estimator $\hat{\mathbf{x}}$, $\underline{\text{not depending on } c_l,\ c_L \text{ and } c_e}$, such that, with high-probability

$$d_\infty(\hat{\mathbf{x}}, \mathbf{x}^*) \leq C_{lLe} \left(\min_{(\mathbf{x},f)\in\mathcal{R}[F,c_l,c_L,c_e]} d_\infty(\mathbf{x}, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}} \right),$$

for some representation $(\mathbf{x}^*, f) \in \mathcal{R}_{\mathbf{\Pi}_n}[F, c_l, c_L, c_e]$, and some constant $C_{lLe} > 0$ depending only on c_e, c_l and c_L. For such an estimator, under the premise that $d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n)$ is small for $(\mathbf{x}^*, f) \in \mathcal{R}_{\mathbf{\Pi}_n}[F, c_l, c_L, c_e]$, we then estimate accurately a representation $(\mathbf{x}^*, f)$ of F in $\mathcal{R}_{\mathbf{\Pi}_n}[F]$.

As a side remark, we notice that combining such a bound with (9), we obtain that, for any $(\mathbf{x}, f) \in \mathcal{R}_{\mathbf{\Pi}_n}[F, c_l, c_L, c_e]$, we have with high-probability

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}) \leq C'_{lLe} \left(\min_{(\mathbf{x}',f')\in\mathcal{R}[F,c_l,c_L,c_e]} d_\infty(\mathbf{x}', \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}} \right).$$

3. Localize-and-Refine algorithm

The overall strategy for estimating a n-tuple of latent positions in $\mathcal{R}_{\mathbf{\Pi}_n}[F]$, is to start with a first estimator $\hat{\mathbf{x}}^{(1)} \in \mathbf{\Pi}_n$ with a control in $d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n d(x_i, y_i)$ distance, and then to refine the estimation of each point x_i^*. In order to avoid complex statistical dependencies between the two steps, we use a sample splitting of the data. We sample $S \subset \{1, \ldots, n\}$ with cardinality $|S| = n_0 = n/4$ uniformly at random, and set $\overline{S} = \{1, \ldots, n\} \setminus S$. The first estimator $\hat{\mathbf{x}}_S^{(1)}$ of $\mathbf{x}_S^* = (x_i^*)_{i\in S}$ is computed on $A_{SS} = [A_{ij}]_{i,j\in S}$, while the refined estimator $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ of $\mathbf{x}_{\overline{S}}^* := (x_i^*)_{i\in\overline{S}}$ takes as input the estimator $\hat{\mathbf{x}}_S^{(1)}$ and the matrix $A_{S\overline{S}}$. This scheme avoids to have some statistical dependence between $\hat{\mathbf{x}}_S^{(1)}$ and $A_{S\overline{S}}$. The estimator $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ provides a localization for points indexed by $\overline{S}$, with an error bound in d_∞-norm. In order to localize all the points, we repeat the process and the final estimator is obtained by carefully merging the estimations. These three steps are precisely described in Sections 3.2–3.4, after the statement of our main results.

---

[2] $\mathcal{R}_{\mathbf{\Pi}_n}[F]$ is well-defined and non-empty because the set of all $\mathbf{x}$ such that, for some f, $(\mathbf{x}, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ is compact.

3.1. Main result

In this section, we consider the following setting for the data.

Setting 1. *Assume that the matrix A of observations is given by (3), with $\mathbf{x}^* \in \mathcal{C}^n$, and $f \in \mathcal{BL}[c_l, c_L, c_e]$, for some $c_e \geq 0$ and $0 < c_l \leq c_L$ (the set of Bi-Lipschitz functions $\mathcal{BL}[c_l, c_L, c_e]$ is introduced in Definition 2.1, page 1594). Assume also that the noise matrix E follows the subGaussian errors assumption (7).*

In this setting, the estimator (21) described in the next subsections fulfills the following risk bound.

Theorem 3.1. *Assume that the data are generated according to the Setting 1 above. Then, there exists a constant $C_{lLe} > 0$ depending only on c_e, c_l and c_L such that, with probability at least $1 - 5/n^2$, there exists a representation $(\mathbf{x}, f)$ in the set $\mathcal{R}_{\mathbf{\Pi}_n}[F, c_l, c_L, c_e]$ defined in (10) such that the estimator (21) fulfills*

$$d_\infty(\hat{\mathbf{x}}, \mathbf{x}) \leq C_{lLe} \left(\min_{(\mathbf{x}', f') \in \mathcal{R}[F, c_l, c_L, c_e]} d_\infty(\mathbf{x}', \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}} \right) , \tag{11}$$

with $\mathcal{R}[F, c_l, c_L, c_e]$ defined by (8).

We emphasize that the estimator (21) has no tuning parameter. In particular, it does not depend on the unknown constants $c_e \geq 0$ and $0 < c_l \leq c_L$. The first term in the right-hand side of (11) can be assimilated to a bias term, which stems from the bias of the estimator (21) towards regular positions on $\mathcal{C}$. The second term in the right-hand side of (11) is a variance-type term. We observe that, if there exists $c_a > 0$ such that

$$\min_{(\mathbf{x}', f') \in \mathcal{R}[F, c_l, c_L, c_e]} d_\infty(\mathbf{x}', \mathbf{\Pi}_n) \leq c_a \sqrt{\frac{\log(n)}{n}} , \tag{12}$$

then $d_\infty(\hat{\mathbf{x}}, \mathbf{x}) = O\left(\sqrt{\log(n)/n}\right)$ with high probability. Such a setting arises for example when the latent positions have been sampled uniformly on the sphere, see Corollary 3.3. This $\sqrt{\log(n)/n}$ rate is shown to be minimax optimal in Section 5.

We also give an explicit control for any representative $(\mathbf{x}, f) \in \mathcal{R}[F, c_l, c_L, c_e]$.

Theorem 3.2. *Under the Setting 1, there exists a constant $C'_{lLe} > 0$ depending only on c_e, c_l and c_L, such that, with probability at least $1 - 5/n^2$, for any $(\mathbf{x}, f) \in \mathcal{R}[F, c_l, c_L, c_e]$, the estimator (21) fulfills*

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}) \leq C'_{lLe} \left(d_\infty(\mathbf{x}, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}} \right) . \tag{13}$$

We emphasize that the above statement holds for *any* representative $(\mathbf{x}, f) \in \mathcal{R}[F, c_l, c_L, c_e]$. Before moving to the description of the estimator (21), let us give two important instantiations of Theorem 3.1 and 3.2.

Latent model with uniform sampling on $\mathcal{C}$ Let us consider the latent model $F_{ij} = f(x_i^*, x_j^*)$ when the positions $x_1^*, \dots, x_n^*$ have been sampled independently and uniformly on $\mathcal{C}$, as in the graphon model. In this case, the Assumption (12) is satisfied with high probability; see Appendix B.5 for a proof. We then derive the next result from Theorem 3.2.

Corollary 3.3. *Assume that the latent positions $x_1^*, \dots, x_n^*$ have been sampled i.i.d. uniformly on $\mathcal{C}$ and that $f \in \mathcal{BL}[c_l, c_L, c_e]$. Then, with probability higher than $1 - 7/n^2$, we have*

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*) \leq C_{lLe} \sqrt{\frac{\log(n)}{n}} \ ,$$

for some constant $C_{lLe} > 0$ depending only on c_e, c_l and c_L.

Toroidal seriation Let us consider the toroidal seriation problem introduced in Example 4 of Section 1.1. In this setting, the set $[n]$ is considered as a torus, endowed with the torus distance $d(i,j) = \min(|j-i|, |n+i-j|)$ for any $1 \leq i \leq j \leq n$, and the matrix F is a pre-toroidal R-matrix. Let $\sigma^* \in \Sigma_n$ be a permutation such that $[F_{\sigma^*(i)\sigma^*(j)}]_{i,j}$ is a toroidal R-matrix. Our goal is estimate σ^* from the noisy observation $A = F + E$. As explained in the introduction, we can recast this problem as a localization problem in a latent space model on the regular grid $\mathcal{C}_n$. Assimilating points on the sphere $\mathcal{C}$ to unit norm complex numbers, we define the vector $\mathbf{x}^* \in \mathbf{\Pi}_n$ by $x_j^* = \exp(\iota 2\pi\sigma^*(j)/n)$, and we define $f_n : \mathcal{C}_n \times \mathcal{C}_n \to \mathbb{R}$ by $f(x_i^*, x_j^*) = F_{ij}$. The problem of estimating σ^* then amounts to estimating $\mathbf{x}^*$ in the latent space model $A_{ij} = f(x_i^*, x_j^*) + E_{ij}$ on $\mathcal{C}_n$. We can apply our estimator (21) and get an estimation $\hat{\mathbf{x}} \in \mathcal{C}^n$. From this estimation, we can derive the map $\hat{\sigma} : [n] \to [n]$ by setting $\hat{\sigma}_i = \lceil n\underline{\hat{x}}_i/2\pi \rceil$ for $i = 1, \dots, n$, where $\underline{\hat{x}}_i \in (0, 2\pi]$ is the argument of $\hat{x}_i$ and $\lceil z \rceil$ is the upper integer part of z. While the map $\hat{\sigma}$ may not be a permutation in Σ_n, it is an estimation of σ^* and we can translate Theorem 3.2 into a ℓ^∞-error between the two.

Corollary 3.4. *Assume that $[F_{\sigma^*(i)\sigma^*(j)}]_{i,j=1,\dots,n}$ fulfills the bi-Lipschitz condition with respect to the torus distance:*

$$\begin{aligned} |F_{\sigma^*(i)\sigma^*(j)} - F_{\sigma^*(i)\sigma^*(k)}| &\leq c_L \frac{2\pi}{n} d(j,k) + \varepsilon_n \\ F_{\sigma^*(i)\sigma^*(k)} - F_{\sigma^*(i)\sigma^*(j)} &\geq c_l \frac{2\pi}{n}\big(d(i,j) - d(i,k)\big) - \varepsilon_n \quad \text{if } d(i,j) \geq d(i,k) \ . \end{aligned}$$

Then, there exists a constant $C_{lLe} > 0$, depending only on c_e, c_l and c_L, such that, with probability at least $1 - 5/n^2$, we have

$$\min_{\tau \in \Gamma_n} \max_{i \in [n]} |\tau \circ \sigma^*(i) - \hat{\sigma}(i)| \leq C_{lLe} \sqrt{n \log(n)} \ , \tag{14}$$

where Γ_n is the subgroup of permutations of $\{1, \dots, n\}$ generated by the circular and reverse permutations.

To prove (14), we extend f_n defined on $\mathcal{C}_n \times \mathcal{C}_n$ to $f : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ belonging to $\mathcal{BL}[c_l, c_L, c_e]$, and apply Theorem 3.2. The minimum over Γ_n in the left-hand side of (14) cannot be avoided, since σ^* is identifiable from F only up to permutations in Γ_n. Furthermore, the $\sqrt{n\log(n)}$ rate for the toroidal seriation problem can be shown to be minimax in the above set-up, by combining Theorem 5.1 page 1608 and the correspondence between the n-tuples $\mathbf{x}^* \in \mathbf{\Pi}_n$ and the permutations σ^* of $[n]$.

In a recent work, Janssen and Smith [25] consider the related seriation problem for R-matrices (Example 3 in the introduction), in a geometric setting where $F_{\sigma_i^*,\sigma_j^*} = g(|i-j|/n)$ for some unknown permutation σ^*, and some unknown function g. Hence, in addition to be a pre-R matrix, F is also a Toeplitz matrix. Under additional assumptions on the squared matrix $(F_{\sigma_i^*,\sigma_j^*})^2$, they establish that an algorithm based on a (thresholded version) of the square matrix of observations A^2, achieves, with high probability, an error bound

$$\min_{\tau\in\Gamma'_n} \max_{i\in[n]} |\tau\circ\sigma^*(i) - \hat{\sigma}(i)| \lesssim \sqrt{n}\,(\log(n))^5,$$

where Γ'_n gathers the identity and the reverse permutations. Their assumptions are not comparable to ours, but their rates are similar (up to log factors). Our results then complement this work, by providing another set of conditions on F, under which the hidden permutation σ^* can be recovered at the rate $\sqrt{n\log(n)}$.

Organization of Section 3 The description of the estimator (21) is organized as follows. The refinement step $\hat{\mathbf{x}}^{(2)}_{\overline{S}}$ is described in Section 3.2. This step can take as input any initial estimator $\hat{\mathbf{x}}^{(1)}_S$ based on A_{SS} and taking values in $\mathbf{\Pi}_{|S|}$. When this initial estimator fulfills with high-probability

$$d_1(\hat{\mathbf{x}}^{(1)}_S, \mathbf{x}_S) \le C_{lLe}\left(n \min_{(\mathbf{x}',f')\in\mathcal{R}[F]} d_\infty(\mathbf{x}'_S, \mathbf{\Pi}_{|S|}) + \sqrt{n\log(n)}\right), \tag{15}$$

for some $(\mathbf{x}, f) \in \mathcal{R}_{\mathbf{\Pi}_n}[F]$, then the refined estimator $\hat{\mathbf{x}}^{(2)}_{\overline{S}}$ is shown to fulfill with high-probability

$$d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}}, \mathbf{x}_{\overline{S}}) \le C'_{lLe}\left(\min_{(\mathbf{x}',f')\in\mathcal{R}[F]} d_\infty(\mathbf{x}'_S, \mathbf{\Pi}_{|S|}) + \sqrt{\frac{\log(n)}{n}}\right).$$

In order to get an estimator satisfying the risk bound (11), we then need an initial estimator $\hat{\mathbf{x}}^{(1)}_S$ fulfilling (15). Such an estimator is provided in Section 3.3. A computationally efficient alternative, based on the spectral decomposition of A is proposed in Section 4. To get an estimator of the whole n-tuple of latent positions, the data splitting is repeated and a final merging step is needed to build $\hat{\mathbf{x}}$. This final step is described in Section 3.4.

3.2. Step 2: refined estimation

We start by describing the refinement step which converts an initial estimator with an error bound in d_1-distance into a refined estimator with an error bound in d_∞-distance. For a subset $S \subset \{1, \ldots, n\}$ of cardinality $|S| = n_0 = n/4$, the refinement step takes as input any initial estimator $\hat{\mathbf{x}}_S^{(1)}$ of $\mathbf{x}_S^*$ based on A_{SS} and taking values in $\mathbf{\Pi}_{n_0}$. It outputs an estimator $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ of $\mathbf{x}_{\overline{S}}^*$.

We denote by $D(z, \hat{\mathbf{x}}_S^{(1)}) = [d(z, \hat{x}_j^{(1)})]_{j \in S}$ the vector of distances between $z \in \mathcal{C}$ and the components of the n_0-tuple $\hat{\mathbf{x}}_S^{(1)}$. The refined estimator $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ is obtained by solving

$$\hat{x}_i^{(2)} \in \operatorname*{argmin}_{z \in \mathcal{C}_{n_0}} \left\langle A_{i,S}, D(z, \hat{\mathbf{x}}_S^{(1)}) \right\rangle, \quad \text{for each } i \in \overline{S}, \tag{16}$$

where $\mathcal{C}_{n_0} = \{1, e^{\iota 2\pi/n_0}, \ldots, e^{\iota 2\pi(n_0-1)/n_0}\}$ is the regular grid of cardinality n_0 on $\mathcal{C}$. The principle underlying the definition (16) is that $d(\hat{x}_i^{(2)}, \hat{x}_j^{(1)})$ should be small when A_{ij} is large, and vice-versa. Hence, for any $x_i^* \in \mathcal{C}_{n_0}$ and any matrix A with A_{ij} decreasing with $d(x_i^*, x_j^*)$, the minimum is achieved in x_i^* when $\hat{\mathbf{x}}_S^{(1)} = \mathbf{x}_S^* \in \mathbf{\Pi}_{n_0}$, see Appendix A for details. Since A is a noisy version of a such a matrix, and since $d(x_i^*, \mathcal{C}_{n_0}) := \min_{y \in \mathcal{C}_{n_0}} d(x_i^*, y) = O(1/n)$, the estimator $\hat{x}_i^{(2)}$ should remain close to x_i^* when $\hat{\mathbf{x}}_S^{(1)}$ is close to $\mathbf{x}_S^*$. The next proposition quantifies this statement, by providing a uniform error bound for $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ in terms of the error bound $d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}_S^*)$ for the initial estimator.

Proposition 3.5. *Let $(\mathbf{x}^*, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ for some $c_e \geq 0$ and $0 < c_l \leq c_L$. Let $S \subset \{1, \ldots, n\}$ be of cardinality $|S| = n_0 = n/4$, and $\hat{\mathbf{x}}_S^{(1)}$ be any initial estimator of $\mathbf{x}_S^*$ based on A_{SS} and taking values in $\mathbf{\Pi}_{n_0}$. Then, there exists a constant C_{lLe} depending only on c_l, c_L and c_e, such that, conditionally on A_{SS}, and for all $Q \in \mathcal{O}$, the estimator (16) fulfills with probability at least $1 - 1/n^2$*

$$d_\infty(\hat{\mathbf{x}}_{\overline{S}}^{(2)}, Q\mathbf{x}_{\overline{S}}^*) \leq C_{lLe} \left(d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \frac{d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}_S^*)}{n} + \sqrt{\frac{\log(n)}{n}} \right) . \tag{17}$$

The right-hand side of (17) is made of three terms. The first one is the uniform approximation error of $\mathbf{x}_S^*$ by $\mathbf{\Pi}_{n_0}$, as defined in (4). It is a bias-type term which stems from the fact that we aim at estimating positions that are almost evenly spaced. The second term accounts for the error of the preliminary estimator $\mathbf{x}_S^{(1)}$ in d_1-distance and the last-one is a variance-type term. The proof of Proposition 3.5 can be found in Appendix B.

The time complexity for computing $\hat{x}_i^{(2)}$ is linear in n_0 and the algorithm can be parallelized for computing $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$. To decrease the time complexity, it is possible to restrict to $z \in \mathcal{C}_{\sqrt{n_0}}$ in (16) instead of $z \in \mathcal{C}_{n_0}$. In that case, the proof of (17) still holds, with different constants.

3.3. Step 1: Initial localization

In view of the above Proposition 3.5, we seek to build an initial estimator $\hat{\mathbf{x}}_S^{(1)}$ fulfilling (15) for some $\mathbf{x}_S = Q\mathbf{x}_S^*$, with $Q \in \mathcal{O}$. Such an estimator can be obtained by solving

$$\hat{\mathbf{x}}_S^{(1)} \in \operatorname*{argmin}_{\mathbf{x}_S \in \mathbf{\Pi}_{n_0}} \langle A_{SS}, D(\mathbf{x}_S) \rangle, \quad \text{with } D(\mathbf{x}_S) = \big[d(x_i, x_j)\big]_{i,j \in S} \ . \tag{18}$$

The estimator $\hat{\mathbf{x}}_S^{(1)}$ is chosen in such a way that the distance $d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)})$ should be small when the signal $F_{ij} = f(x_i^*, x_j^*)$ is large, and conversely, it should be large when F_{ij} is small. To grasp the principle underlying the definition (18), let us look at the noiseless geometric affine setting, where the observations are $A_{ij} = 1 - \alpha d(x_i^*, x_j^*)$, and where the positions $\mathbf{x}_S^*$ are evenly spread, i.e. $\mathbf{x}_S^* \in \mathbf{\Pi}_{n_0}$. Then, one readily checks that

$$\operatorname*{argmin}_{\mathbf{x}_S \in \mathbf{\Pi}_{n_0}} \langle A_{SS}, D(\mathbf{x}_S) \rangle = \operatorname*{argmax}_{\mathbf{x}_S \in \mathbf{\Pi}_{n_0}} \langle D(\mathbf{x}_S^*), D(\mathbf{x}_S) \rangle \ ,$$

whose maximum is achieved at all $\mathbf{x}_S = Q\mathbf{x}_S^*$ with Q any orthogonal transformation preserving $\mathbf{\Pi}_{n_0}$. In other words, the estimator (18) exactly recovers –up to distance preserving transformations– the positions in this ideal setting.

The next proposition establishes a d_1-bound with the flavor of (15) for the estimator (18).

Proposition 3.6. *Let $(\mathbf{x}^*, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ and let $S \subset \{1, \ldots, n\}$ be a subset of cardinality $n_0 = n/4$. Then, there exists a constant C_{lLe} depending only on c_l, c_L and c_e, such that, the estimator $\hat{\mathbf{x}}_S^{(1)}$ defined by (18) satisfies*

$$\min_{Q \in \mathcal{O}} d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}_S^*) \le C_{lLe} \left(n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n \log n} \right) \ , \tag{19}$$

with probability higher than $1 - 1/n^2$.

To prove Proposition 3.6 (in Appendix B.3), we first establish that $\|D(\mathbf{x}_S^*) - D(\hat{\mathbf{x}}_S^{(1)})\|_2$ is small, meaning that the distances between the estimated positions $\{\hat{x}_i^{(1)} : i \in S\}$ are close to the distances between the true positions $\{x_i^* : i \in S\}$. Then, relying on a recent result on matrix perturbation from [1], we deduce that $\|\hat{\mathbf{x}}_S^{(1)} - Q\mathbf{x}_S^*\|_2$ is small, where $Q \in \mathcal{O}$ is a distance preserving transformation and where we consider here $\hat{\mathbf{x}}_S^{(1)}$ and $Q\mathbf{x}_S^*$ as $2 \times n_0$ matrices. The bound (19) then follows by connecting the Euclidean distance in $\mathbb{R}^2$ to the d_1-distance.

From a computational point of view, the minimization problem (18) is an instance of the Quadratic Assignment Problem which is known to be NP-Hard and even hard to approximate [31, 35]. In section 4, we propose a computationally efficient alternative to (18), and we provide theoretical guarantees for this alternative under additional model assumptions.

3.4. Final merging step

For a given subset $S \subset \{1, \ldots, n\}$ of cardinality $n_0 = n/4$, combining the initial estimator (18) with the refined localisation (16), we get an estimator $\hat{\mathbf{x}}_S^{(2)}$ with an error bound on $d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}}, Q\mathbf{x}^*_{\overline{S}})$ for some orthogonal transformation $Q \in \mathcal{O}$. In order to get an estimation of all the latent positions, we repeat the process by sampling $S' \subset \overline{S}$ of cardinality $n_0 = n/4$ and by computing $\hat{\mathbf{x}}^{(2')}_{\overline{S'}}$ with (18) and (16). We then get an estimator with an error bound on $d_\infty(\hat{\mathbf{x}}^{(2')}_{\overline{S'}}, Q'\mathbf{x}^*_{\overline{S'}})$ for some orthogonal transformation $Q' \in \mathcal{O}$. In order to get a final estimator $\hat{\mathbf{x}}$, we still have to deal with the fact that we may have $Q \neq Q'$, and hence the trivial merge $\hat{\mathbf{x}} = (\hat{\mathbf{x}}^{(2)}_{\overline{S}}, \hat{\mathbf{x}}^{(2')}_{S})$ may not be a good one. Hence, we need to synchronize the estimators $\hat{\mathbf{x}}^{(2)}_{\overline{S}}$ and $\hat{\mathbf{x}}^{(2')}_{\overline{S'}}$. This synchronization is obtained by solving

$$\widehat{Q} \in \operatorname*{argmin}_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S'}}, Q\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S'}}), \tag{20}$$

and by defining the final estimator as $\hat{\mathbf{x}} = (\hat{\mathbf{x}}^{(2)}_{\overline{S}}, \widehat{Q}\hat{\mathbf{x}}^{(2')}_{S})$. Putting pieces together, we then get the following estimation procedure.

Localize-and-Refine procedure

Input: Observations matrix A.

A) Localization of $3n/4$ points

1. Pick uniformly at random a subset $S \subset [n]$ of cardinality $|S| = n/4$.
2. Compute $\hat{\mathbf{x}}^{(2)}_{\overline{S}}$ by solving

$$\begin{aligned} \hat{\mathbf{x}}^{(1)}_{S} \quad &\in \quad \operatorname*{argmin}_{\mathbf{x}_S \in \mathbf{\Pi}_{n_0}} \langle A_{SS}, D(\mathbf{x}_S) \rangle \,, \\ \hat{x}^{(2)}_{i} \quad &\in \quad \operatorname*{argmin}_{z \in \mathcal{C}_{n_0}} \langle A_{i,S}, D(z, \hat{\mathbf{x}}^{(1)}_{S}) \rangle, \quad \text{for } i \in \overline{S}\,, \end{aligned}$$

with $D(\mathbf{x}_S) = \big[d(x_i, x_j)\big]_{i,j\in S}$ and $D(z, \mathbf{x}_S) = \big[d(z, x_j)\big]_{j \in S}$.

B) Localization of $3n/4$ (other) points

1. Pick uniformly at random a subset $S' \subset \overline{S}$ of size $|S'| = n/4$.
2. Compute $\hat{\mathbf{x}}^{(2')}_{\overline{S'}}$ by solving

$$\begin{aligned} \hat{\mathbf{x}}^{(1')}_{S'} \quad &\in \quad \operatorname*{argmin}_{\mathbf{x}_{S'} \in \mathbf{\Pi}_{n_0}} \langle A_{S'S'}, D(\mathbf{x}_{S'}) \rangle, \\ \hat{x}^{(2')}_{i} \quad &\in \quad \operatorname*{argmin}_{z \in \mathcal{C}_{n_0}} \langle A_{i,S'}, D(z, \hat{\mathbf{x}}^{(1')}_{S'}) \rangle, \quad \text{for } i \in \overline{S}'\,, \end{aligned}$$

with $D(\mathbf{x}_{S'}) = \big[d(x_i, x_j)\big]_{i,j\in S'}$ and $D(z, \mathbf{x}_{S'}) = \big[d(z, x_j)\big]_{j \in S'}$.

C) Merging the two localizations

1. Compute $\widehat{Q}$ by solving (20).
2. Output: $\hat{\mathbf{x}} \in \mathcal{C}^n$ defined by

$$\hat{x}_j = \begin{cases} \hat{x}^{(2)}_j & \text{if } j \in \overline{S}\ , \\ \widehat{Q}\hat{x}^{(2')}_j & \text{if } j \in S\ . \end{cases} \tag{21}$$

The minimization problem (20) can be solved efficiently. For example, we can observe that the minimum is achieved at some $\hat{Q} \in \mathcal{O}$ preserving $\mathcal{C}_{n_0}$, and since there are at most $2n_0$ such orthogonal transformations, we can enumerate them. We refer to Section 6 for details.

4. Spectral localization in the geometric latent model

The computation of the initial localization $\hat{\mathbf{x}}_S^{(1)}$ requires to minimize (18) over $\mathbf{\Pi}_{n_0}$, which is an instance of the Quadratic Assignment Problem (QAP), which is known to be NP-hard and hard to approximate [31, 35]. A spectral relaxation of the QAP has been shown to be successful for reordering a pre (toroidal) R-matrix [3, 33], and hence for solving the noiseless seriation problem for R-matrices. This vanilla spectral algorithm proposed in [3] takes as input any symmetric matrix $M \in \mathbb{R}^{N\times N}$ and output N points in $\mathbb{R}^2$.

Vanilla Spectral Algorithm (VSA)

Input: symmetric matrix $M \in \mathbb{R}^{N\times N}$ with eigenvalues $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{N-1}$.

Compute: two orthonormal eigenvectors $\hat{u}, \hat{v} \in \mathbb{R}^N$ associated with the second and third eigenvalues $\hat{\lambda}_1$ and $\hat{\lambda}_2$ of M

Output:

$$\hat{\mathbf{x}}^{\mathrm{VSA}} = \left(\sqrt{\frac{N}{2}}(\hat{u}_1, \hat{v}_1)^T, \ldots, \sqrt{\frac{N}{2}}(\hat{u}_N, \hat{v}_N)^T\right) \in \mathbb{R}^{2\times N} \tag{22}$$

In this section, we adapt this vanilla spectral algorithm in order to get a computationally efficient initial estimator $\tilde{\mathbf{x}}_S^{(1)}$. In Section 4.1, we describe the estimator $\tilde{\mathbf{x}}_S^{(1)}$ and provide some error bounds in d_∞-distance for the Localize-and-Refine algorithm based on $\tilde{\mathbf{x}}_S^{(1)}$. The main difference compared to Section 3 is that our theory is limited to cases where the function f is *geometric*, that is,

$$f(x, y) = g(d(x, y)) \quad \text{for all } x, y \in \mathcal{C}, \tag{23}$$

for some $g : [0, \pi] \to [0, 1]$. In Section 4.2, we complement this result by providing an error bound in ℓ^1-norm for the vanilla spectral algorithm (VSA) in the geometric case.

4.1. Spectral localization algorithm

We observe that the output $\hat{\mathbf{x}}_S^{\mathrm{VSA}}$ of the vanilla spectral algorithm applied to A_{SS} does not belong to $\mathbf{\Pi}_{n_0}$, and even not to $\mathcal{C}^{n_0}$. Hence, we need an additional approximation step in order to get an estimator $\tilde{\mathbf{x}}_S^{(1)}$ that can be plugged in our Localize-and-Refine procedure (16). In the description of the algorithm below, we identify points on the circle $\mathcal{C}$ to unit norm complex numbers. Besides, for such a point z, we write $\|z\|_1$ for its ℓ^1 norm in $\mathbb{R}^2$.

The next theorem provides an error bound in d_∞-distance for the Localize-and-Refine procedure (21), when we replace $\hat{\mathbf{x}}_S^{(1)}$ by $\tilde{\mathbf{x}}_S^{(1)}$. This bound involves

```
Spectral Localization (LS)
Input: a subset $S = \{i_1, \ldots, i_{n_0}\} \subset [n]$ with $i_1 < \ldots < i_{n_0}$, and data matrix $A_{SS}$.

Vanilla spectral localization: compute
$\hat{\mathbf{x}}_S^{\mathrm{VSA}} := [\hat{x}_{i_\ell}^{\mathrm{VSA}}]_{l \in [n_0]} = \mathrm{VSA}(A_{SS})$

Uniform Approximation (UA) in $\mathbf{\Pi}_{n_0}$:
    1. Set $z_\ell = \hat{x}_{i_\ell}^{\mathrm{VSA}} / \|\hat{x}_{i_\ell}^{\mathrm{VSA}}\|_2$, for $\ell = 1, \ldots, n_0$.
    2. Pick any permutation $\sigma$ such that $z_{\sigma(1)}, \ldots, z_{\sigma(n_0)}$ is in trigonometric order.
    3. Set $\tilde{x}_{i_{\sigma(\ell)}}^{(1)} = e^{\iota \frac{2\pi(\widehat{k}+\ell)}{n_0}}$,   where   $\widehat{k} \in \mathrm{argmin}_{k \in [n_0]} \sum_{\ell=1}^{n_0} \left\| e^{\iota \frac{2\pi(k+\ell)}{n_0}} - z_{\sigma(\ell)} \right\|_1$.
Output: $\tilde{\mathbf{x}}_S^{(1)} := [\tilde{x}_{i_\ell}^{(1)}]_{l \in [n_0]} \in \mathbf{\Pi}_{n_0}$.
```

the two spectral gaps $\Delta_1 = \lambda_0^* - \lambda_1^*$ and $\Delta_2 = \lambda_2^* - \lambda_3^*$, where $\lambda_0^* \geq \ldots \geq \lambda_{n-1}^*$ denote the eigenvalues of the signal matrix F.

Theorem 4.1. *Let $n \geq 16$, $c_a, c_b, c_e > 0$, and $0 < c_l \leq c_L$. Let $(\mathbf{x}^*, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ with f a geometric function $f = g \circ d$. Assume that $\mathbf{x}^*$ fulfills*

$$d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq c_a \sqrt{\frac{\log(n)}{n}} \;, \tag{24}$$

and that the spectral gaps satisfy $\Delta_1 \wedge \Delta_2 \geq c_b n$. Then, there exists a constant $C_{lLeab} > 0$ depending only on c_e, c_l, c_L, c_a and c_b, such that, with probability at least $1 - 9/n^2$, the spectral Localize-and-Refine procedure (21) with $\hat{\mathbf{x}}_S^{(1)}$ and $\hat{\mathbf{x}}_{S'}^{(1')}$ replaced by $\tilde{\mathbf{x}}_S^{(1)}$ and $\tilde{\mathbf{x}}_{S'}^{(1')}$ satisfies the uniform bound

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*) \leq C_{lLeab} \sqrt{\frac{\log(n)}{n}} \;.$$

Similarly as in Theorem 3.1 and 3.2, we estimate the latent positions at the optimal $\sqrt{\log(n)/n}$ rate in d_∞-distance, but under the additional assumptions that f is a geometric function (23) and F fulfills the spectral gap condition $\Delta_1 \wedge \Delta_2 \geq c_b n$. The proof of Theorem 4.1 is given in Appendix D. The proof mainly relies on controlling the ℓ^1-norm between $\hat{\mathbf{x}}^{\mathrm{VSA}}$ and $Q\mathbf{x}^*$. This result, which has its own interest, is presented in Proposition 4.4, in Section 4.2. Below, we exhibit two cases where the spectral gap condition $\Delta_1 \wedge \Delta_2 \geq c_b n$ holds.

Example: Geometric model with Fourier gaps The eigenvalues of F are closely related to the discrete Fourier transform of g, so that we can bound the spectral gaps Δ_1 and Δ_2 in terms of these Fourier coefficients. More precisely, the function f is given by $f(x, y) = g(d(x, y))$, with g defined on $[0, \pi]$. One can extend g to $[0, 2\pi)$ by taking $g(x) = g(2\pi - x)$ for any $x \in (\pi, 2\pi)$. Then, for any integer n, the discrete Fourier transform of $\left\{g(j\frac{2\pi}{n}) : j = 0, \ldots, n-1\right\}$ is defined by

$$\mathcal{F}_{k,n}(g) = \sum_{j=0}^{n-1} g\left(j\frac{2\pi}{n}\right)\cos\left(j\frac{2\pi k}{n}\right), \quad \text{for } k = 0, \ldots, n-1. \tag{25}$$

The following lemma bounds the spectral gaps Δ_1 and Δ_2 in terms of the gaps between the Fourier coefficients.

Lemma 4.2. *Let $c_a, c_b, c_e > 0$, and $0 < c_l \leq c_L$. Let $(\mathbf{x}^*, f) \in \mathcal{R}[F, c_l, c_L, c_e]$ with f a geometric function $f = g \circ d$, and $\mathbf{x}^*$ fulfilling (24).*

Let us set $\Phi_1 = \mathcal{F}_{0,n}(g) - \mathcal{F}_{1,n}(g)$ and $\Phi_2 = \min_{j=2,\ldots,\lfloor n/2 \rfloor} \mathcal{F}_{1,n}(g) - \mathcal{F}_{j,n}(g)$. Then, there exists a constant $C_{lLea} > 0$, depending only on c_e, c_l, c_L and c_a, such that

$$|\Delta_1 - \Phi_1| \vee |\Delta_2 - \Phi_2| \leq C_{lLea}\sqrt{n\log(n)}\ .$$

Hence, Theorem 4.1 still holds when we replace the gap condition $\Delta_1 \wedge \Delta_2 \geq c_b n$ by the condition $\Phi_1 \wedge \Phi_2 \geq c_b n$. So, when the first discrete Fourier coefficients of g are well separated from the other coefficients, the spectral version of the Localize-and-Refine algorithm estimates, in polynomial time, the latent positions at the optimal $\sqrt{\log(n)/n}$ rate in d_∞-distance. Below, we give an example where the Fourier coefficients can be explicitly computed and where Φ_1, Φ_2 are proportional to n.

Example: Affine geometric model As a simple instantiation of Theorem 4.1 and Lemma 4.2, let us consider the geometric function $f(x, y) = 1 - d(x, y)/(2\pi)$. The corresponding univariate function $g(z) = 1 - z/(2\pi)$ is affine and its discrete Fourier coefficients can be computed explicitly in terms of trigonometric functions. In Appendix D.6, we prove that $\Phi_1 \wedge \Phi_2 \geq c_b n$ for some numerical constant $c_b > 0$. We then get the next corollary of Theorem 4.1.

Corollary 4.3. *Let f be the function defined as $f(x, y) = 1 - d(x, y)/(2\pi)$. Assume that the latent positions $\mathbf{x}^* \in \mathcal{C}^n$ fulfill (24). Then, there exist constants C_a and C'_a depending only on c_a such that for all $n \geq C'_a$, with probability higher than $1 - 9/n^2$, the spectral Localize-and-Refine procedure (21), with $\hat{\mathbf{x}}_S^{(1)}$ and $\hat{\mathbf{x}}_{S'}^{(1')}$ replaced by $\tilde{\mathbf{x}}_S^{(1)}$ and $\tilde{\mathbf{x}}_{S'}^{(1')}$, satisfies the uniform bound*

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*) \leq C_a \sqrt{\frac{\log(n)}{n}}\ .$$

Theorem 5.1 in the next section shows that this estimation rate is optimal.

4.2. ℓ^1-bound for the vanilla spectral algorithm

As a byproduct of our analysis, we provide an ℓ^1-bound for the estimation of the latent positions with the vanilla spectral algorithm (VSA), in the geometric latent model. Recanati et al. [33] have already shown that VSA succeeds to recover the hidden permutation in the noiseless seriation problem with R-matrices. We extend their work to the geometric latent model on $\mathcal{C}$.

Starting from the noisy observation $A = F + E$ with $F_{ij} = g \circ d(x_i^*, x_j^*)$, we apply VSA to the whole matrix A and get an estimation $\hat{\mathbf{x}}^{\text{VSA}} \in \mathbb{R}^{2\times n}$ of $\mathbf{x}^*$. The next proposition provides a bound in terms of the ℓ^1-distance

$$\|\hat{\mathbf{x}}^{\text{VSA}} - \mathbf{x}\|_1 := \sum_{j=1}^{n}\sum_{i=1}^{2} \left|\hat{x}_{ij}^{\text{VSA}} - x_{ij}\right| ,$$

and in terms of the spectral gaps $\Delta_1 = \lambda_0^* - \lambda_1^*$ and $\Delta_2 = \lambda_2^* - \lambda_3^*$, where $\lambda_0^* \geq \ldots \geq \lambda_{n_0-1}^*$ are the eigenvalues of the signal matrix F.

Proposition 4.4. *Let $n \geq 4$, and let $f = g \circ d \in \mathcal{BL}[c_l, c_L, c_e]$ be a bi-Lipschitz geometric function. Assume that the latent positions $\mathbf{x}^*$ fulfill the Assumption (24), with $c_a > 0$. Then, there exist two constants C_{lLea} and C'_{lLea}, depending only on c_l, c_L, c_e and c_a, such that, with probability at least $1-1/n^2$, the vanilla spectral estimator $\hat{\mathbf{x}}^{VSA}$ satisfies*

$$\begin{aligned}\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}^{VSA} - Q\mathbf{x}^*\|_1 &\leq C_{lLea}\,\frac{n\sqrt{n\log(n)}}{(\Delta_1 \wedge \Delta_2) \vee 1}\\ &\leq C_{lLea}\,\frac{n\sqrt{n\log(n)}}{\left[(\Phi_1 \wedge \Phi_2) - C'_{lLea}\sqrt{n\log(n)}\right] \vee 1} .\end{aligned}$$

Proposition 4.4 is proved in Appendix D.7. It provides an ℓ^1-localization bound depending on the spectral gap $\Delta_1 \wedge \Delta_2$ of the signal matrix F. Since there are only n positions to be estimated in the bounded space $\mathcal{C}$, this bound is uninformative when the spectral gaps $\Delta_1 \wedge \Delta_2$ are smaller than $\sqrt{n\log(n)}$. Conversely, when the spectral gaps are of the order of n, we get an ℓ^1-bound of the desired scaling $\sqrt{n\log(n)}$.

Proposition 4.4 is based on the fact that the signal matrix F is well approximated by a circulant and circular-R matrix, which benefits from nice spectral properties, see Appendix D.4. This type of R-matrices was already studied in [33] to derive some error bounds on the reconstruction of positions – see Proposition D in [33]. Here, Proposition 4.4 extends their result by providing some explicit bounds in the stochastic setting and also by considering some more general signals F, which are not assumed to be an exact circulant and circular R-matrix.

5. Minimax lower bound

In this section, we prove that the $\sqrt{\log(n)/n}$ rate in Theorem 3.1 is minimax optimal. Let us consider the observation model $A = F + E$, where we assume that the entries $\{A_{ij} : i < j\}$ follow independent Bernoulli distributions with parameters $f(x_i^*, x_j^*)$. We focus on this specific case of sub-Gaussian distributions in the lower bound, as we have in mind random graph applications. We emphasize that the same lower-bound holds for Gaussian noise.

To prove the lower bound, we consider the simpler setting where f_0 is known to the statistician, and is an affine function of d,

$$f_0(x,y) = (3/4) - d(x,y)/(4\pi)\,, \quad \text{for all } x,y \in \mathcal{C}.$$

This function f_0 corresponds to a geometric latent model as discussed in the introduction, and it satisfies the bi-Lipschitz assumption (5- 6) for $c_e = 0$ and $c_l = c_L = (4\pi)^{-1}$. In this simple scenario, the latent positions are identifiable up to the orthogonal transformations in $\mathcal{O}$, so we derive a lower bound in terms of the quasi-metric $\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*)$. Recall that $\mathbb{P}_{(\mathbf{x}^*,f_0)}$ denotes the distribution of A with representation $(\mathbf{x}^*, f_0)$.

Theorem 5.1. *There exist two positive constants C, C' such that for any $n \geq C'$, we have the lower bound*

$$\inf_{\hat{\mathbf{x}}} \sup_{\mathbf{x}^*\in\mathbf{\Pi}_n} \mathbb{P}_{(\mathbf{x}^*,f_0)}\left[\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*) \geq C\sqrt{\frac{\log(n)}{n}}\right] \geq \frac{1}{2}\ ,$$

where the infimum holds over all $\sigma(A)$-measurable functions $\hat{\mathbf{x}}$.

The proof of the Theorem 5.1 is given in Appendix C.3. The lower bound is written over the collection of n-tuples $\mathbf{x}^* \in \mathbf{\Pi}_n$, which is a subclass of the class considered in our upper bounds (since all $\mathbf{x}^* \in \mathbf{\Pi}_n$ satisfy the condition (12) for any $c_a \geq 0$). The lower bound matches the upper bound in Theorem 3.2 up to some multiplicative constants. Therefore, it implies the optimality of the $\sqrt{\log(n)/n}$ estimation rate of our estimator (in the minimax sense). The fact that the lower bound holds even for a known function entails that the rate $\sqrt{\log(n)/n}$ is not driven by the (absence of) knowledge of the affinity function in our setting. Moreover, since the affine function f_0 satisfies the bi-Lipschitz assumption (5–6) for $c_e = 0$, i.e. $f_0 \in \mathcal{BL}[(4\pi)^{-1}, (4\pi)^{-1}, 0]$, this entails that the rate $\sqrt{\log(n)/n}$ is not due to the slack $\varepsilon_n = c_e\sqrt{\log(n)/n}$ in the bi-Lipschitz assumption. In fact, we precisely allow this slack $c_e\sqrt{\log(n)/n}$ in (5–6) because this generalization does not worsen the estimation rate compared to pure bi-Lipschitz functions ($c_e = 0$). Finally, since the set of n-tuples $\mathbf{x}^* \in \mathbf{\Pi}_n$ is in correspondence with the set of permutations σ^* of $[n]$, Theorem 5.1 ensures that the bound (14) is rate-optimal for the bi-Lipschitz seriation problem.

6. Numerical experiments

6.1. Optimal rate

In Figure (1), we study the ratio $r = \frac{\min_{Q\in\mathcal{Q}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*)}{v_{opt}}$ of the maximum error of the Localize-and-Refine algorithm $\hat{\mathbf{x}}$ (without data splitting) and the optimal rate $v_{opt} = \sqrt{\log(n)/n}$. For each sample size $n = 100, 200, 300, 400$, a dot represents the average of 50 ratios $r_1, \ldots, r_{50}$ obtained on independent data sets $A^{(1)}, \ldots, A^{(50)}$. Each data matrix $A^{(j)}$, $j \in [50]$, has been generated as in the

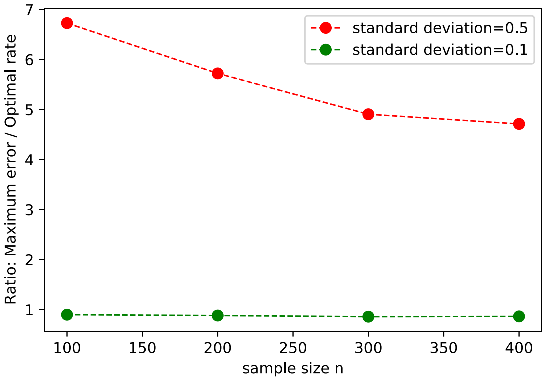


FIG 1. *Ratio of the maximum error of the Localize-and-Refine algorithm and the optimal rate* $\sqrt{\log(n)/n}$*, with noise level* $\sigma = 0.1$ *(green) and* $\sigma = 0.5$ *(red).*

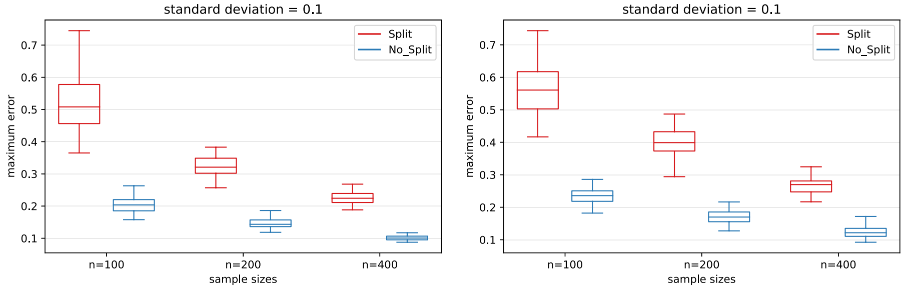
model (3), with the three following specifications. The latent points $x_1^*, \ldots, x_n^*$ are sampled independently and uniformly on $\mathcal{C}$. The affinity function is the affine geometric function $f(x,y) = 1 - d(x,y)/(2\pi)$. The entries E_{ij}, $1 \leq i < j \leq n$, of the noise matrix are independent Gaussian random variables, with a standard deviation that is either equal to 0.1 (green curve) or to 0.5 (red curve).

One can observe in Figure (1) that the (averaged) ratio for $\sigma = 0.1$ is (approximately) constant and equal to 1, while for $\sigma = 0.5$ it decreases from 7 to 5. In other words, the maximum error of the Localize-and-Refine algorithm follows a $\sqrt{\log(n)/n}$ rate, up to a multiplicative constant $C \in [\frac{1}{2}, 8]$, for sample sizes $n \geq 100$. This corroborates the conclusion of our theoretical findings (upper bound of Corollary 3.3 and lower bound of Theorem 5.1) that the Localize-and-Refine algorithm achieves the optimal $\sqrt{\log(n)/n}$ rate up to a multiplicative constant C that is bounded away from zero and bounded from above. An interesting question (for future research) would be to understand the dependencies of C in the problem parameters. Figure (1) indeed shows that C behaves differently when $\sigma = 0.1$ or $\sigma = 0.5$, and that C varies with n.

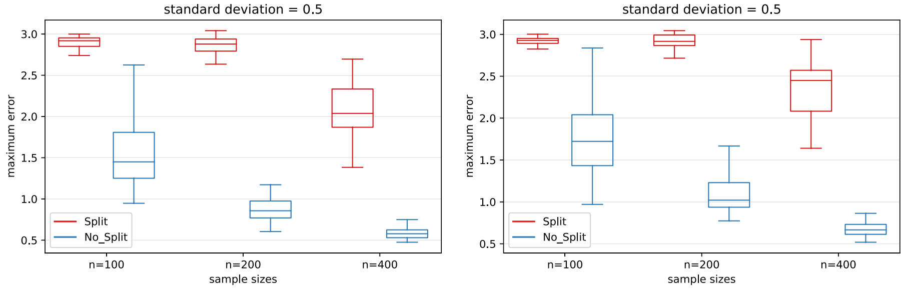
6.2. Usefulness of data splitting? of refined estimation step 2?

In this section, we investigate two questions relative to the empirical performance of the Localize-and-Refine algorithm with the initial localization $\hat{\mathbf{x}}_S^{(1)}$ given by the spectral output $\tilde{\mathbf{x}}_S^{(1)}$ (defined page 1605): (i) Is the data splitting useful in practice? (ii) Does the refined estimation (step 2) empirically improve the initial localization (step 1)?

In each Figure 2 and 3, we compare two algorithms, presenting boxplots of their localization errors in d_∞-distance. Each boxplot represents the distribution of 50 errors made on 50 samplings of the data matrices A. Each data matrix A is generated as in the model (3), with the three following spec-



(a) affine signal (b) logit signal



(c) affine signal (d) logit signal

FIG 2. *Localization error in d_∞-distance for the Localize-and-Refine algorithm with data-splitting (red) and without data-splitting (blue). Top line: noise with standard-deviation sd= 0.1. Bottom line: noise with sd=0.5. Left: affine geometric affinity function $f(x,y) = 1 - d(x,y)/(2\pi)$. Right: Logit geometric affinity function $f(x,y) = \exp[-d(x,y)]/\left(1+\exp[-d(x,y)]\right)$.*

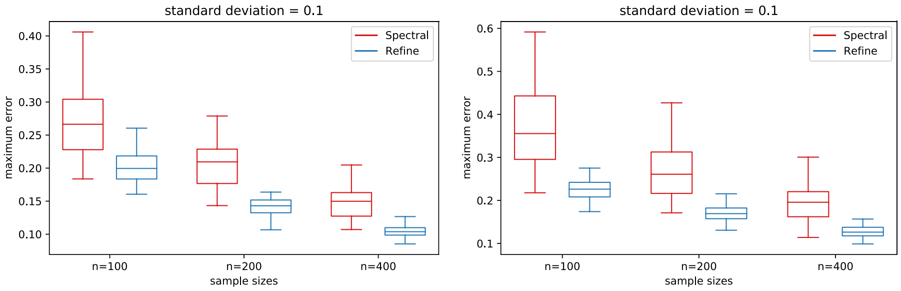
ifications. The latent points $x_1^*, \dots, x_n^*$ are sampled independently and uniformly on $\mathcal{C}$. For the affinity function, we choose either the affine geometric function $f(x,y) = 1 - d(x,y)/(2\pi)$, or the logit geometric function $f(x,y) = \exp[-d(x,y)]/\left(1+\exp[-d(x,y)]\right)$. In the noise matrix, the entries E_{ij}, $1 \le i < j \le n$, are independent Gaussian random variables, with a standard deviation that is either equal to 0.1 (top line) or 0.5 (bottom line). The same protocol is used in Figure 4, except that we measure the localization error in d_1-distance, instead of d_∞-distance.

Question (i): We use a data splitting scheme in the Localize-and-Refine algorithm in order to ensure independence between the data used in the two steps. This independence was convenient to prove theoretical guarantees (as Theorem 4.1). Yet, data splitting makes the initial localization run on a $(n/4) \times (n/4)$ data matrix, instead of the whole $n \times n$ matrix, which is expected to enlarge the variance of this initial localization by a factor 4. So, one can wonder whether the splitting is necessary and useful in practice. To answer this question, we illustrate in Figure 2 the difference between the performances of the Localize-and-Refine algorithm and the homologous procedure without data splitting (the former is plotted in red, the latter in blue). One can observe that the d_∞-localization error is much smaller for the procedure without splitting. A plausible explanation for the good performances without data-splitting is that the statistical dependence between the steps 1 and 2 of the algorithm is negligible for large n, rendering the data splitting useless. Indeed, in the no-splitting version of the algorithm, step 1 uses $O(n^2)$ observations to release a first localization $\tilde{\mathbf{x}}^{(1)}$ of the n positions, then step 2 refines the estimation of a position x_i using n observations, which only represents a fraction $O(1/n)$ of the observations used in step 1. Hence, the dependence between $\tilde{\mathbf{x}}^{(1)}$ in step 1 and the n observations in step 2 could be sufficiently small to not require a data splitting. Accordingly, we recommend the version of the Localize-and-Refine algorithm without data-spliting for practical use. As a future direction of research, it would be interesting to investigate the theoretical performance of the algorithm without data splitting, in order to bridge the gap between the theory and the practice.

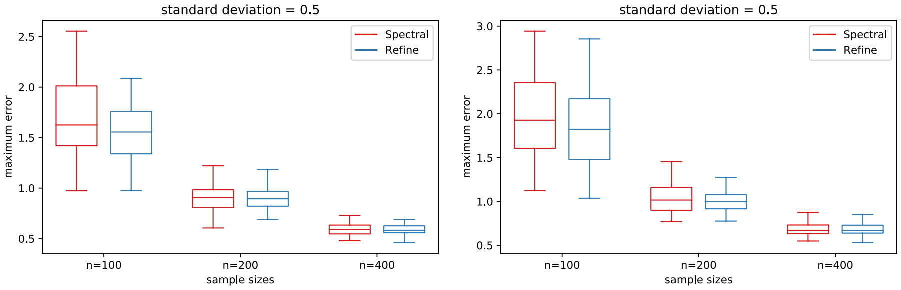
Question (ii): The strategy of the Localize-and-Refine algorithm is to get an initial localization with controlled d_1-error and then to refine the localization in order to ensure a control in the d_∞-metric. A natural question is whether the refinement step 2 improves the initial localization obtained by the Spectral algorithm in step 1. We investigate numerically this question in Figure 3, by comparing the d_∞-error of the Spectral Localization procedure (plotted in red) and of the Localize-and-Refine algorithm without data splitting (in blue). One can observe contrasting results, depending on the standard-deviation of the noise. When the standard-deviation is 0.5 (bottom line), the second step offers no significant improvement in the d_∞-localization error. Conversely, when the standard-deviation is 0.1 (top line), the d_∞-error is significantly improved by the refinement step. This suggests that, to be useful, the refinement step requires a precise enough initial localization. We complement Figure 3 with Figure 4, which displays the errors in d_1-distance (scaled by $1/n$ for a better comparison), instead of the d_∞-distance, though this loss function is not our main concern in this paper. In Figure 4, we observe a behavior in the (d_1/n)-metric very similar to the behavior in the d_∞-metric, displayed in Figure 3. In the light of the numerical performance of the Spectral Localization in Figure 3, an interesting open question is wether we can prove theoretical guarantees on the d_∞-localization error of this procedure.

7. Discussion

Relying on observations of pairwise affinities in a latent space model, we studied the problem of uniformly localizing positions $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ on the unit sphere $\mathcal{C} \subset \mathbb{R}^2$. Under bi-Lipschitz assumptions on the affinity function, we established the rate $\sqrt{\log(n)/n}$ for the uniform localization of balanced n-tuples $\mathbf{x}^* \in \mathbf{\Pi}_n$. We also proved that non-trivial estimation error is still possible when the latent points do not form a balanced n-tuple ($\mathbf{x}^* \notin \mathbf{\Pi}_n$) to the price of an additional bias $d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n)$. This bias remains small compared to



(a) affine signal (b) logit signal



(c) affine signal (d) logit signal

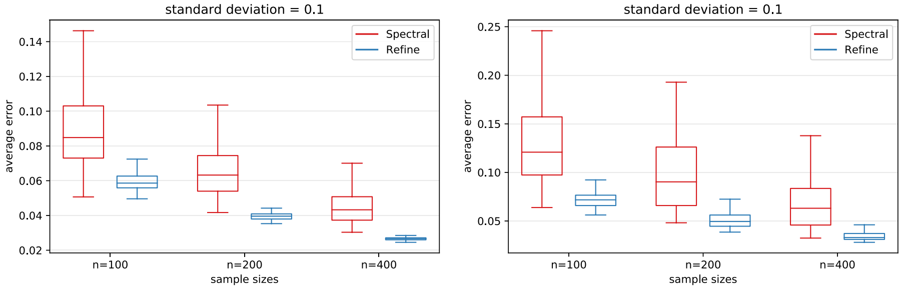
Fig 3. *Localization error in d_∞-distance for the Spectral algorithm (red) and the Localize-and-Refine algorithm (blue). Top line: noise with sd=0.1. Bottom line: noise with sd=0.5. Left: affine geometric affinity function. Right: Logit geometric affinity function.*

the $\sqrt{\log(n)/n}$ rate when the points have been sampled uniformly at random on $\mathcal{C}$.

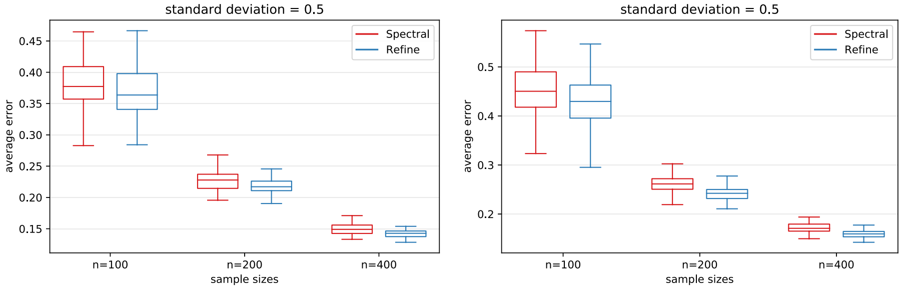
We also analyzed a spectral embedding alternative in Section 4, which benefits from a polynomial-time complexity. When the function f is geometric and when the associated Fourier coefficients are suitably separated, this spectral method achieves the optimal rate $\sqrt{\log(n)/n}$ for uniform localization. Yet, the spectral embedding takes advantage of the structure of Toeplitz R-matrix. Since this structure disappears in the general case of bi-Lipschitz functions, there is no apparent reason for the spectral algorithm to work over the whole class of bi-Lipschitz functions.

As our non-polynomial-time algorithm is based on an instance of the Quadratic Assignment Problem, which is known to be NP Hard and even hard to approximate, the existence of polynomial-time algorithms achieving the $\sqrt{\log(n)/n}$ rate over the whole class of bi-Lipschitz functions remains an open question.

The latent positions are not identifiable when f is unknown, and our main hypothesis is that there exists a representation $(\mathbf{x}, f)$ with $\mathbf{x}$ close to Π_n and f bi-Lipschitz. We use as reference the regular distribution Π_n, since regular and uniform distributions are the ones that appear in classical models like



(a) affine signal (b) logit signal



(c) affine signal (d) logit signal

FIG 4. *Localization error in* $\left(\frac{1}{n}d_1\right)$*-distance for the Spectral algorithm (red) and the Localize-and-Refine algorithm (blue). Top line: noise with sd=0.1. Bottom line: noise with sd=0.5. Left: affine geometric affinity function. Right: Logit geometric affinity function.*

graphon, f-random graphs, or statistical seriation. Our algorithms builds on this hypothesis, and consequently the bias $\min_{(\mathbf{x},f)\in\mathcal{R}(c_l,c_L,c_e)} d_\infty(\mathbf{x},\Pi_n)$ appears in our bounds, where the minimum is over the set $\mathcal{R}(c_l,c_L,c_e)$ of bi-Lipschitz representatives $(\mathbf{x},f)$. This minimum leaves room to handle situations where the latent positions do not match the regular grid Π_n but are only more or less evenly spread. For instance, the minimal bias is zero for some representations $(\mathbf{x},f)$, with $\mathbf{x}$ as far apart from Π_n as $d_\infty(\mathbf{x},\Pi_n) \geq \pi/8$ (Proposition 2.2). Yet, there are many practical situations where the affinity matrix is clustered, that we cannot handle. In the case where the affinity matrix is clustered (f bi-Lipschitz, but the x_i are clustered), the problem becomes a clustering problem, rather than a seriation problem, and our algorithms are not suited for clustering data. The question of handling simultaneously clustering and seriation is very interesting, but it is beyond the scope of this paper.

In this manuscript, we focused our attention to symmetric pairwise affinity functions f. However, other one-dimensional localization models such as Bradley-Terry model or more generally ranking problems, do not satisfy the symmetry assumption. Still, we hope that our general two-step approach can leverage other structural assumptions. In ranking, a natural counterpart of our model (3) is the so-called SST model introduced by [36], which is defined as follows. We observe $A_{ij} = f(x_i^*, x_j^*) + E_{ij}$ where the function $f : (x,y) \in [0,1] \times [0,1] \mapsto f(x,y) \in [0,1]$ is non-decreasing with respect to x and non-increasing with respect to y and satisfies the skew symmetry assumption, that is $f(x,y) = 1 - f(y,x)$. In this setting, $f(x_i^*, x_j^*)$ stands for the probability that player i wins a game against player j. Note that the latent space is now $[0,1]$ and not the torus $\mathcal{C}$ anymore. Although our methodology does not apply verbatim, we could adapt the Localize and Refine procedure for the latent space $[0,1]$. To exploit the bi-isotonic and skew-symmetric assumptions, the refinement estimator of (16) could for instance be replaced by

$$\hat{x}_i^{(2)} \in \operatorname*{argmax}_{z \in G_{n_0}} \left\langle (A_{i,S} - \frac{1}{2}), z - \hat{\mathbf{x}}_S^{(1)} \right\rangle, \quad \text{for each } i \in \overline{S} ,$$

where G_{n_0} stands for the regular grid $\{\frac{1}{n_0}, \frac{2}{n_0}, \dots, 1\}$ and $\hat{\mathbf{x}}_S^{(1)}$ is a suitable first-step estimator. In comparison to (16), $D(z, \hat{\mathbf{x}}_S^{(1)})$ is replaced by $z - \hat{\mathbf{x}}_S^{(1)}$. We expect that, with a suitable initialization $\hat{\mathbf{x}}_S^{(1)}$ and under bi-Lipschitz assumptions, the resulting procedure achieves near-optimal localization rates. This is an interesting direction for future research.

Appendix A: Some intuition on our analysis

To get some intuition on the rationale behind our analysis, we single out the next lemma, which is a cornerstone of the analysis at least in the simplified situation where $c_e = 0$ and where the latent positions belong to $\mathbf{\Pi}_n$. Then, we discuss some consequences in simplified versions of our work.

Lemma A.1. *Let $\alpha, \varepsilon \geq 0$ be two non-negative constants, and let $(a_j)_{j=1,\dots,p}$ and $(d_j)_{j=1,\dots,p}$ be two sequences fulfilling*

$$a_1 \geq a_2 \geq \dots \geq a_p, \quad d_1 \leq d_2 \leq \dots \leq d_p, \quad \textit{and} \quad a_j - a_{j+1} \geq \alpha(d_{j+1} - d_j) - \varepsilon ,$$

for $j = 1, \dots, p-1$. Then, for any permutation $\sigma : [p] \to [p]$ we have

$$\sum_{j=1}^{p} a_j (d_{\sigma(j)} - d_j) \geq \frac{\alpha}{2} \sum_{j=1}^{p} (d_j - d_{\sigma(j)})^2 - \varepsilon \sum_{j=1}^{p} j(d_j - d_{\sigma(j)}). \tag{26}$$

Proof of Lemma A.1. Let us set the notation $D_j = d_1 + \dots + d_j$, and $d_j^\sigma = d_{\sigma(j)}$, and $D_j^\sigma = d_1^\sigma + \dots + d_j^\sigma$ for $j = 1, \dots, p$. Since d is non-decreasing, and since σ is a permutation of $[p]$, we have

$$D_j^\sigma \geq D_j, \quad \text{for } j = 1, \dots, p, \quad \text{and} \quad D_p^\sigma = D_p \ .$$

Writing $d_j^\sigma = D_j^\sigma - D_{j-1}^\sigma$ and rearranging the sums, we get

$$\sum_{j=1}^{p} a_j (d_j^\sigma - d_j) = \sum_{j=1}^{p} a_j \big((D_j^\sigma - D_{j-1}^\sigma) - (D_j - D_{j-1}) \big)$$

$$
\begin{aligned}
&= \underbrace{a_p \left(D^\sigma_p - D_p\right)}_{=0} + \sum_{j=1}^{p-1} \underbrace{(a_j - a_{j+1})}_{\geq \alpha(d_{j+1}-d_j)-\varepsilon} \underbrace{(D^\sigma_j - D_j)}_{\geq 0} \\
&\geq \sum_{j=1}^{p-1} [\alpha(d_{j+1} - d_j) - \varepsilon(j+1-j)](D^\sigma_j - D_j) \\
&= \sum_{j=1}^{p} (\alpha d_j - \varepsilon j)(d_j - d^\sigma_j) \\
&= \frac{\alpha}{2} \sum_{j=1}^{p} (d_j - d^\sigma_j)^2 - \varepsilon \sum_{j=1}^{p} j(d_j - d^\sigma_j) \ ,
\end{aligned}
$$

where we used Abel transformation in the penultimate line. The proof of Lemma A.1 is complete. □

Let us discuss some immediate consequences of the above lemma for our problem. Let us consider the case where the entries A_{ij} of the matrix A decrease with $d(x^*_i, x^*_j)$ for some $\mathbf{x}^* \in \mathbf{\Pi}_n$. For a fixed i, let τ be a permutation of $[n]$ such that $\left\{d(x^*_i, x^*_{\tau(j)}) : j = 1, \dots, n\right\}$ is ranked in increasing order. Let us set $a_j = A_{i\tau(j)}$ and $d_j = d(x^*_i, x^*_{\tau(j)})$. Since the entries A_{ij} decrease with $d(x^*_i, x^*_j)$, the sequences $(a_j)_{j=1,\dots,n}$ and $(d_j)_{j=1,\dots,n}$ fulfill the conditions of Lemma A.1 with $\alpha = \varepsilon = 0$. Let us pick $k \in [n]$ and let us denote by σ_k the permutation of $[n]$ such that $d(x^*_k, x^*_{\tau(j)}) = d(x^*_i, x^*_{\tau(\sigma_k(j))}) = d_{\sigma_k(j)}$ – this is possible because $\mathbf{x}^* \in \mathbf{\Pi}_n$. We notice that $\sigma_i = Id$. Then, Lemma A.1 ensures that

$$
\sum_{j=1}^{n} A_{i\tau(j)} d(x^*_k, x^*_{\tau(j)}) = \sum_{j=1}^{n} A_{i\tau(j)} d(x^*_i, x^*_{\tau(\sigma_k(j))}) \geq \sum_{j=1}^{n} A_{i\tau(j)} d(x^*_i, x^*_{\tau(j)}) \ ,
$$

so that

$$
x^*_i \in \operatorname*{argmin}_{z \in \mathcal{C}_n} \sum_{j=1}^{n} A_{ij} d(z, x^*_j) \ .
$$

This justifies that the criterion underlying the refined estimator (16) is able to recover the true latent position x^*_j at least in an idealized setting where the observations are noiseless, the entries of A_{ij} are decreasing with $d(x^*_i, x^*_j)$, and the true latent positions x^*_j are plugged in (16) instead of the initial estimator $\widehat{x}^{(1)}$.

When, in addition, we have a lower Lipschitz condition

$$
A_{i\tau(j)} - A_{i\tau(j+1)} \geq c_l \left(d(x^*_i, x^*_{\tau(j+1)}) - d(x^*_i, x^*_{\tau(j)}) \right) \ ,
$$

then, applying Lemma A.1, we can lower bound the difference

$$
\sum_{j=1}^{n} A_{ij} d(x^*_k, x^*_j) - \sum_{j=1}^{n} A_{ij} d(x^*_i, x^*_j) \geq \frac{c_l}{2} \sum_{j=1}^{n} \left(d(x^*_i, x^*_j) - d(x^*_k, x^*_j) \right)^2 \ .
$$

In particular, we observe that, for all $z \in \mathcal{C}_n$,

$$\sum_{j=1}^{n} A_{ij} d(z, x_j^*) \geq \sum_{j=1}^{n} A_{ij} d(x_i^*, x_j^*) + \frac{c_l}{2} \sum_{j=1}^{n} \left(d(x_i^*, x_j^*) - d(z, x_j^*)\right)^2 ,$$

so the sum $\sum_j A_{ij} d(z, x_j^*)$ locally increases, when z moves away from x_i^*. In the general case, where $c_e > 0$ and the observations are noisy, the criterion does not satisfy a simple local quadratic lower bound and we need to rely on finer arguments than Lemma A.1 – see e.g. the proofs Lemma B.3 and B.16.

Finally, we sketch here the proof of the second result of Proposition 2.3, in the specific case where $c_e = 0$. Consider any two representations $(\mathbf{x}, f)$ and $(\mathbf{x}, f')$ in $\mathcal{R}[F, c_l, c_L, 0]$ with $\mathbf{x}$, $\mathbf{x}' \in \mathbf{\Pi}_n$. Since both $\mathbf{x}$ and $\mathbf{x}'$ belong to $\mathbf{\Pi}_n$, this implies that, for any fixed i, the vectors $(d(x_i, x_j))_j$ and $(d(x_i', x_j'))_j$ are equal, up to a permutation of the entries. As $c_e = 0$, the lower Lipschitz condition (6) ensures that $(F_{i,j})_{j=1,\dots,n}$ is decreasing both with respect to $d(x_i, x_j)$ and $d(x_i', x_j')$. As a consequence, we have $d(x_i, x_j) = d(x_i', x_j')$ for any i, j in $[n]$. We now show that this implies that $\mathbf{x} = Q\mathbf{x}'$ for some $Q \in \mathcal{O}$. Denote σ the permutation of $[n]$ such that $\sigma(1) = 1$ and the arguments $\underline{x}_j$ satisfy $\underline{x}_{\sigma(i+1)} = \underline{x}_{\sigma(i)} + 2\pi/n$. As a consequence, we have $d(\mathbf{x}'_{\sigma(i)}, \mathbf{x}'_{\sigma(i+1)}) = d(\mathbf{x}_{\sigma(i)}, \mathbf{x}_{\sigma(i+1)}) = 2\pi/n$. This implies that either $\underline{x'}_{\sigma(i+1)} = \underline{x'}_{\sigma(i)} + 2\pi/n$ for all i, or $\underline{x'}_{\sigma(i+1)} = \underline{x'}_{\sigma(i)} - 2\pi/n$ for all i. In the former case, one easily sees that $\mathbf{x} = Q\mathbf{x}'$, where Q is the rotation satisfying $x_1 = Qx_1'$, whereas in the latter case, we have $\mathbf{x} = Q\mathbf{x}'$, where Q is the reflection satisfying $x_1 = Qx_1'$.

Appendix B: Proofs of main results

Recall that $n = 4n_0$. Given an orthogonal transformation $Q \in \mathcal{O}$, we define the d_1-loss relative to Q as

$$\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) := \frac{d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}_S^*)}{n_0} \quad . \tag{27}$$

Before proving Proposition 3.5, we study, as a warm-up, the simpler situation where all the latent positions x_i^* are elements of the regular grid $\mathcal{C}_{n_0}$ and where the vector $\mathbf{x}_S^*$ (composed of n_0 coordinates of $\mathbf{x}^*$) belongs to $\mathbf{\Pi}_{n_0}$. In this case, $d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) = 0$.

Lemma B.1. *In addition of the assumptions listed in Proposition 3.5, we assume that $\mathbf{x}^* \in \mathcal{C}_{n_0}^n$ and $\mathbf{x}_S^* \in \mathbf{\Pi}_{n_0}$. Then for any $Q \in \mathcal{O}$ and any $i \in \overline{S}$, the estimator (16) satisfies the following bound*

$$d_\infty(\hat{x}_i^{(2)}, Qx_i^*) \leq C_{l,L,e} \left(\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + \sqrt{\frac{\log(n)}{n}}\right) \quad ,$$

with probability at least $1 - 1/n^3$.

Taking a union bound over the $n - n_0$ indices $i \in \overline{S}$, we

$$d_\infty(\hat{\mathbf{x}}_{\overline{S}}, Q\mathbf{x}^*_{\overline{S}}) \leq C_{l,L,e}\left(\mu_{\hat{\mathbf{x}}^{(1)}_S,\mathbf{x}^*_S}(Q) + \sqrt{\frac{\log(n)}{n}}\right) \ ,$$

with probability higher than $1 - 1/n^2$. This is exactly the conclusion of Proposition 3.5 in the special case where $\mathbf{x}^* \in \mathcal{C}^n_{n_0}$ and $\mathbf{x}^*_S \in \mathbf{\Pi}_{n_0}$. The proof of Proposition 3.5 for general $\mathbf{x}^*$ follows the same scheme as that of Lemma B.1, but also requires some slight refinements. We first prove Lemma B.1 before turning to the general case.

B.1. Proof of Lemma B.1

First, we claim that it suffices to restrict our attention to transformations $Q \in \mathcal{O}$ that let $\mathcal{C}_{n_0}$ invariant. Indeed, for general Q, there exists an orthogonal transformation Q', letting $\mathcal{C}_{n_0}$ invariant, and such that $\max_{z\in\mathcal{C}_{n_0}} d(Qz, Q'z) \lesssim 1/n_0$. Replacing Q' by Q in the statement of Lemma B.1 only entails an additional term of order $1/n_0$ which is negligible compared to the term $\sqrt{\log(n)/n}$.

Let $i \in \overline{S}$. In the two next lemmas, we bound

$$L_i := \langle F_{i,S},\ D(Q^{-1}\hat{x}^{(2)}_i, \mathbf{x}^*_S) - D(x^*_i, \mathbf{x}^*_S)\rangle \tag{28}$$

from above and below. We recall that $F_{i,S}$ is the vector $(f(x^*_i, x^*_j)$ for $j \in S$.

Lemma B.2. *With probability at least $1 - 1/n^3$, we have*

$$L_i \leq C_{L,e}d(\hat{x}^{(2)}_i, Qx^*_i)\left(n\mu_{\hat{\mathbf{x}}^{(1)}_S,\mathbf{x}^*_S}(Q) + \sqrt{n\log(n)}\right) \ ,$$

for some constant $C_{L,e} > 0$.

Lemma B.3. *We have*

$$L_i \geq Cnd(\hat{x}^{(2)}_i, Qx^*_i)\Big(c_l d(\hat{x}^{(2)}_i, Qx^*_i) - \varepsilon_n\Big) - \frac{c_e^3}{\pi c_l^2}\sqrt{\frac{\log^3(n)}{n}} - 2\frac{c_e}{c_l}\sqrt{\frac{\log(n)}{n}} \ ,$$

for some numerical constant $C > 0$ and all n larger than quantity $C_{l,e}$.

These two lemmas imply that, for n large enough and

$$d(\hat{x}^{(2)}_i, Qx^*_i) \geq 2C'_{l,e}\sqrt{\log(n)/n} \ ,$$

with $C'_{l,e}$ large enough, we have

$$C'_{l,e}nd^2(\hat{x}^{(2)}_i, Qx^*_i) \leq L_i \leq C_{L,e}d(\hat{x}^{(2)}_i, Qx^*_i)\left(n\mu_{\hat{\mathbf{x}}^{(1)}_S,\mathbf{x}^*_S}(Q) + \sqrt{n\log(n)}\right) \ .$$

We conclude that the error bound $d(\hat{x}^{(2)}_i, Qx^*_i) \leq C_{l,L,e}[\mu_{\hat{\mathbf{x}}^{(1)}_S,\mathbf{x}^*_S}(Q) + \sqrt{\log(n)/n}]$ holds with probability at least $1 - 1/n^3$.

B.1.1. Proof of Lemma B.2

Since $|S| = n_0$, we can assume that $S = [n_0]$ for the ease of exposition. Let $i \in \overline{S}$. First, we decompose L_i as follows

$$L_i = \sum_{j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(Q^{-1}\hat{x}_i^{(2)}, x_j^*) - d(x_i^*, x_j^*)\Big).$$

The regular grid $\mathcal{C}_{n_0}$ is invariant by Q, and $\mathbf{x}_S^*$ belongs to $\mathbf{\Pi}_{n_0}$. Besides, $\hat{\mathbf{x}}_S^{(1)}$ belongs to $\mathbf{\Pi}_{n_0}$. As a consequence, we have $\{x_j^*;\, j \in [n_0]\} = \mathcal{C}_{n_0} = \{Q^{-1}\hat{x}_j^{(1)};\, j \in [n_0]\}$. Hence, we can reorder the sum in L_i as follows

$$L_i = \sum_{j=1}^{n_0} f(x_i^*, Q^{-1}\hat{x}_j^{(1)})\Big(d(Q^{-1}\hat{x}_i^{(2)}, Q^{-1}\hat{x}_j^{(1)}) - d(x_i^*, Q^{-1}\hat{x}_j^{(1)})\Big) \ .$$

To alleviate the notation, we write $\hat{z}_i := Q^{-1}\hat{x}_i^{(2)}$ and $z_j^{(1)} := Q^{-1}\hat{x}_j^{(1)}$ so that

$$L_i = \sum_{j=1}^{n_0} f(x_i^*, z_j^{(1)})\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) \ .$$

Lemma B.4. *We have*

$$\sum_{j=1}^{n_0} \Big(f(x_i^*, z_j^{(1)}) - f(x_i^*, x_j^*)\Big)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big)$$
$$\le C_{L,e} d(\hat{x}_i^{(2)}, Qx_i^*)\left[n\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + \sqrt{n\log(n)}\right] \ .$$

Gathering this lemma with the definition of L_i leads us to

$$L_i \le C_{L,e} d(\hat{x}_i^{(2)}, Qx_i^*)\left[n\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + \sqrt{n\log(n))}\right] + \sum_{j=1}^{n_0} f(x_i^*, x_j^*)\left(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\right) \ . \tag{29}$$

The orthogonal transformation Q preserves the distances, hence the last term of (29) is equal to

$$\sum_{j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) = \langle F_{i,S},\, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \ . \tag{30}$$

To handle this term, we come back to the definition (16) of $\hat{x}_i^{(2)}$. Since $Qx_i^* \in \mathcal{C}_{n_0}$, we have

$$\langle A_{i,S}, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle \le \langle A_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \ .$$

This yields

$$\langle F_{i,S}, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \le \langle E_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle \ .$$

The right hand-side $\langle E_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle$ depends on $\hat{x}_i^{(2)}$ which belongs to $\mathcal{C}_{n_0}$. This is why we simultaneously control the expression $\langle E_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(z, \hat{\mathbf{x}}_S^{(1)})\rangle$ for all $z \in \mathcal{C}_{n_0}$. This expression is distributed as a mean zero sub-Gaussian random variable with norm at most $C\|D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(z, \hat{\mathbf{x}}_S^{(1)})\|_2$. Applying a union bound over all $z \in \mathcal{C}_{n_0}$ leads us to

$$\langle E_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle \leq C\sqrt{\log(n_0)}\|D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\|_2$$

with probability higher than $1 - 1/n^3$. Invoking the triangular inequality for the distance d, we deduce that $\|D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\|_2 \leq d(\hat{x}_i^{(2)}, Qx_i^*)\sqrt{n_0}$. It follows that, with probability at least $1 - 1/n^3$,

$$\langle F_{i,S},\ D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \leq Cd(\hat{x}_i^{(2)}, Qx_i^*)\sqrt{n_0\log(n_0)}\ .$$

Gathering this bound with (29) and (30) concludes the proof. □

Proof of Lemma B.4. The first bi-Lipschitz condition (5) ensures that

$$|f(x_i^*, z_j^{(1)}) - f(x_i^*, x_j^*)| \leq c_L d(z_j^{(1)}, x_j^*) + \varepsilon_n\ .$$

By triangular inequality, we also have $|d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})| \leq d(\hat{z}_i, x_i^*)$ so that

$$\sum_{j=1}^{n_0}\Big(f(x_i^*, z_j^{(1)}) - f(x_i^*, x_j^*)\Big)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) \\ \leq d(\hat{z}_i, x_i^*)\sum_{j=1}^{n_0}(c_L d(z_j^{(1)}, x_j^*) + \varepsilon_n)\ .$$

We have $d(z_j^{(1)}, x_j^*) = d(\hat{x}_j^{(1)}, Qx_j^*)$ since Q is an orthogonal transformation. Hence, we obtain

$$\sum_{j=1}^{n_0}\Big(f(x_i^*, z_j^{(1)}) - f(x_i^*, x_j^*)\Big)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) \\ \leq d(\hat{x}_i^{(2)}, Qx_i^*)\left(c_L n_0 \mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + c_e\sqrt{n\log(n)}\right)\ .$$

□

B.1.2. Proof of Lemma B.3

An interval $I = [a, b]$ denotes the set of points lying between a and b in the one-dimensional torus $\mathbb{R}/(2\pi)$, when following the trigonometric direction from a to b. The length of I is denoted by $|I|$. For any point x in the sphere $\mathcal{C}$, its argument in $[0, 2\pi)$ is denoted by $\underline{x}$.

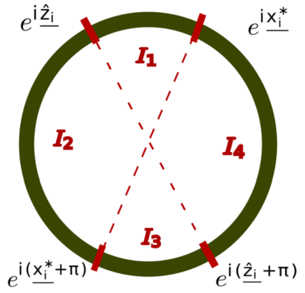

$e^{i\underline{\hat{z}}_i}$
$e^{i\underline{x}_i^*}$
I_1
I_2
I_4
I_3
$e^{i(\underline{x}_i^*+\pi)}$
$e^{i(\underline{\hat{z}}_i+\pi)}$

FIG 5. *Partition of* $[n_0]$ *in* I_1–I_4.

Since $|S| = n_0$, we can assume that $S = [n_0]$ for the ease of exposition. Let $i \in \overline{S}$ and denote $\hat{z}_i := Q^{-1}\hat{x}_i^{(2)}$. Since $d(x_i^*, \hat{z}_i) \leq \pi$, we can assume without loss of generality that the arguments $\underline{x}_i^* = 0$ and $\underline{\hat{z}}_i \in (0, \pi]$, so that we have the equality $d(x_i^*, \hat{z}_i) = |\underline{x}_i^* - \underline{\hat{z}}_i|$. If $\hat{z}_i = x_i^*$, Lemma B.3 is trivial. We therefore assume in the following that $\underline{\hat{z}}_i \in (0, \pi]$. Below, we introduce a partition of $[n_0]$ according to the relative positions of x_j^*, x_i^* and $\hat{z}_i$. This partition is depicted in Figure 5.

$$I_1 = \{j \in [n_0] : \underline{x}_j^* \in [\underline{x}_i^*, \underline{\hat{z}}_i)\} \ ; \qquad I_2 = \{j \in [n_0] : \underline{x}_j^* \in [\underline{\hat{z}}_i, \underline{x}_i^* + \pi)\} \ ;$$
$$I_3 = \{j \in [n_0] : \underline{x}_j^* \in [\underline{x}_i^* + \pi, \underline{\hat{z}}_i + \pi)\} \ ; \qquad I_4 = \{j \in [n_0] : \underline{x}_j^* \in [\underline{\hat{z}}_i + \pi, \underline{x}_i^*)\} \ .$$

Although I_s stands for a subset of indices, with a slight abuse of notation, we still write $|I_s|$ for the length of the corresponding interval in $\mathbb{R}/(2\pi)$. For instance, $|I_1| := |\underline{x}_i^* - \underline{\hat{z}}_i|$.

We decompose L according to this partition of indices $L_i = L_i^{(1)} + L_i^{(2)} + L_i^{(3)} + L_i^{(4)}$, where $L_i^{(s)}$ is the restriction of L_i to the set I_s. In particular, if $\underline{\hat{z}}_i = \pi$, then the intervals I_2 and I_4 are empty, and $L_i^{(2)} = L_i^{(4)} = 0$.

Next, we heavily rely on the fact that the elements of $\mathbf{x}_S^*$ are evenly spaced on the sphere, that is $\{x_i^* : i \in [n_0]\} = \mathcal{C}_{n_0}$ which holds true since we have assumed $\mathbf{x}_S^* \in \mathbf{\Pi}_{n_0}$. Using the symmetry of the set $\mathcal{C}_{n_0}$, we establish below that the sums $L_i^{(2)}$ and $L_i^{(4)}$ nearly compensate so that $L_i^{(2)} + L_i^{(4)}$ admits a positive lower bound.

Lemma B.5. *We have*

$$L_i^{(2)} + L_i^{(4)} \geq \frac{n_0|I_4|}{2\pi} c_l d^2(\hat{x}_i^{(2)}, Qx_i^*) - n_0 d(\hat{x}_i^{(2)}, Qx_i^*)\varepsilon_n \ .$$

As for $L_i^{(1)}$ (resp. $L_i^{(3)}$), we rely on the symmetry of I_1 (resp. I_3) around the point of $\mathcal{C}$ whose argument is $(\underline{x}_i^* + \underline{\hat{z}}_i)/2$ (resp. $((\underline{x}_i^* + \underline{\hat{z}}_i)/2) + \pi$).

Lemma B.6. *For some numerical constant* $C > 0$, *we have*

$$L_i^{(1)} + L_i^{(3)} \geq Cn\Big(\frac{|I_1|}{4} - c_l^{-1}\varepsilon_n\Big)\frac{d(\hat{x}_i^{(2)}, Qx_i^*)}{2}\Big(c_l \frac{d(\hat{x}_i^{(2)}, Qx_i^*)}{2} - \varepsilon_n\Big)$$

$$- (\pi c_l^2)^{-1} c_e^3 \sqrt{\log^3(n)/n} - 2\frac{c_e}{c_l}\sqrt{\frac{\log(n)}{n}} \ .$$

By definition, $|I_1| + |I_4| = \pi$, which yields the desired bound

$$L_i \geq Cnd(\hat{x}_i^{(2)}, Qx_i^*)\Big(c_l d(\hat{x}_i^{(2)}, Qx_i^*) - \varepsilon_n\Big) - \frac{c_e^3}{\pi c_l^2}\sqrt{\frac{\log^3(n)}{n}} - 2\frac{c_e}{c_l}\sqrt{\frac{\log(n)}{n}} \ . \qquad \square$$

Proof of Lemma B.5. In Figure 5, we can see that the difference $d(\hat{z}_i, x_j^*) - d(x_i^*, x_j^*)$ is equal to $-d(\hat{z}_i, x_i^*)$ for all $j \in I_2$, whereas it is equal to $d(\hat{z}_i, x_i^*)$ for $j \in I_4$. Thus, we obtain

$$\begin{aligned} L_i^{(2)} &= \sum_{j \in I_2} f(x_i^*, x_j^*)\Big(d(\hat{z}_i, x_j^*) - d(x_i^*, x_j^*)\Big) = -d(\hat{z}_i, x_i^*) \sum_{j \in I_2} f(x_i^*, x_j^*) \ ; \\ L_i^{(4)} &= \sum_{j \in I_4} f(x_i^*, x_j^*)\Big(d(\hat{z}_i, x_j^*) - d(x_i^*, x_j^*)\Big) = d(\hat{z}_i, x_i^*) \sum_{j \in I_4} f(x_i^*, x_j^*) \ . \end{aligned}$$

Let ϕ denote the reflection with respect to the line going through the two points of $\mathcal{C}$ of arguments

$$a = \frac{\underline{x}_i^* + \underline{\hat{z}}_i}{2} \quad \text{and} \quad b = \frac{(\underline{x}_i^* + \pi) + (\underline{\hat{z}}_i + \pi)}{2}.$$

As can be checked in Figure 5, for any $l \in I_2$, we have $\phi(x_j^*) = x_l^*$ for some j in I_4. Hence,

$$L_i^{(2)} + L_i^{(4)} = d(\hat{z}_i, x_i^*) \sum_{j \in I_4} \big(f(x_i^*, x_j^*) - f(x_i^*, \phi(x_j^*))\big) \ .$$

To lower bound the difference in the sum, we invoke the bi-Lipschitz condition (6), which gives

$$f(x_i^*, x_j^*) - f(x_i^*, \phi(x_j^*)) \geq c_l(d(x_i^*, \phi(x_j^*)) - d(x_i^*, x_j^*)) - \varepsilon_n \ ,$$

since x_j^* is closer to x_i^* than $\phi(x_j^*)$ – see again Figure 5. Also, we can check from Figure 5 that $d(x_i^*, \phi(x_j^*)) - d(x_i^*, x_j^*) = d(\hat{z}_i, x_i^*)$ for all $j \in I_4$. Since $\mathcal{C}_{n_0}$ is evenly spaced, the number of indices j in I_4 is larger than $n_0|I_4|/(2\pi)$. This leads us to

$$L_i^{(2)} + L_i^{(4)} \geq \frac{n_0|I_4|}{2\pi} d(\hat{z}_i, x_i^*) c_l d(\hat{z}_i, x_i^*) - n_0 d(\hat{z}_i, x_i^*)\varepsilon_n \ .$$

Since $d(\hat{z}_i, x_i^*) = d(\hat{x}_i^{(2)}, Qx_i^*)$, this concludes the proof. $\square$

Proof of Lemma B.6. From Figure 5, we see that, for all $j \in I_1$,

$$d(\hat{z}_i, x_j^*) - d(x_i^*, x_j^*) = |\underline{\hat{z}}_i - \underline{x}_j^*| - |\underline{x}_i^* - \underline{x}_j^*| = \underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x}_j^* \ .$$

For $\alpha \in (0,1)$, write $I_1^{(\alpha)}$ the sub-interval of I_1 defined as

$$I_1^{(\alpha)} = \left\{ j \in [n_0] : \underline{x}_j^* \in [\underline{x}_i^*, (1-\alpha)\underline{x}_i^* + \alpha \underline{\hat{z}}_i) \right\} \quad .$$

In particular, for all $j \in I_1^{(1/2)}$, the above expression leads us to

$$d(\hat{z}_i, \phi(x_j^*)) - d(x_i^*, \phi(x_j^*)) = -\big(\underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x_j^*}\big) \quad ,$$

where ϕ is the symmetry introduced in the proof of Lemma B.5. Hence, the terms with $j \in I_1^{(1/2)}$ partially compensate with the terms with j outside $I_1^{(1/2)}$.

$$\begin{aligned} L_i^{(1)} &= \sum_{j \in I_1} f(x_i^*, x_j^*) \left[d(\hat{z}_i, x_j^*) - d(x_i^*, x_j^*) \right] \\ &= \sum_{j \in I_1^{(1/2)}} \left[f(x_i^*, x_j^*) - f(x_i^*, \phi(x_j^*)) \right] \left[\underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x}_j^* \right] \quad . \end{aligned}$$

For any $j \in I_1^{(1/2)}$, we have $d(x_i^*, \phi(x_j^*)) \geq d(x_i^*, x_j^*)$. As a consequence, it follows from the bi-Lipschitz condition (6) that

$$f(x_i^*, x_j^*) - f(x_i^*, \phi(x_j^*)) \geq c_l \left[d(x_i^*, \phi(x_j^*)) - d(x_i^*, x_j^*) \right] - \varepsilon_n \quad .$$

Since $d(x_i^*, \phi(x_j^*)) - d(x_i^*, x_j^*) = |\underline{\phi(x_j^*)} - \underline{x}_j^*|$ for all $j \in I_1^{(1/2)}$, we get

$$L_i^{(1)} \geq \sum_{j \in I_1^{(1/2)}} (c_l |\underline{\phi(x_j^*)} - \underline{x}_j^*| - \varepsilon_n)(\underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x}_j^*) \quad . \tag{31}$$

To control (31), we split the interval $I_1^{(1/2)}$ according to the sign of the term $(c_l|\underline{\phi(x_j^*)} - \underline{x}_j^*| - \varepsilon_n)$. That is, we write $I_1^{(1/2)} = I_1^{(1/2)-} \cup I_1^{(1/2)+}$ where $I_1^{(1/2)-}$ is the set of indices j such that $c_l|\underline{\phi(x_j^*)} - \underline{x}_j^*| < \varepsilon_n$.

Claim B.7. *We have*

$$\begin{aligned} &\sum_{j \in I_1^{(1/2)-}} (c_l |\underline{\phi(x_j^*)} - \underline{x}_j^*| - \varepsilon_n)(\underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x}_j^*) \\ &\qquad\qquad \geq -(c_l^2 8\pi)^{-1} c_e^3 \sqrt{\log^3(n)/n} - \frac{c_e}{c_l}\sqrt{\frac{\log(n)}{n}} \quad . \end{aligned}$$

Claim B.8. *For some numerical constant $C > 0$, we have*

$$\begin{aligned} &\sum_{j \in I_1^{(1/2)+}} (c_l |\underline{\phi(x_j^*)} - \underline{x}_j^*| - \varepsilon_n)(\underline{\hat{z}}_i + \underline{x}_i^* - 2\underline{x}_j^*) \\ &\qquad\qquad \geq Cn\Big(\frac{|I_1|}{4} - c_l^{-1}\varepsilon_n\Big)\frac{|\underline{\hat{z}}_i - \underline{x}_i^*|}{2}\Big(c_l \frac{|\underline{\hat{z}}_i - \underline{x}_i^*|}{2} - \varepsilon_n\Big) \quad . \end{aligned}$$

Gathering these two claims leads us to

$$L_i^{(1)} \geq Cn\big(\frac{|I_1|}{4} - c_l^{-1}\varepsilon_n\big)\frac{|\underline{\hat{z}}_i - \underline{x}_i^*|}{2}\big(c_l\frac{|\underline{\hat{z}}_i - \underline{x}_i^*|}{2} - \varepsilon_n\big) \\ - (c_l^2 8\pi)^{-1}c_e^3\sqrt{\log^3(n)/n} - \frac{c_e}{c_l}\sqrt{\frac{\log(n)}{n}} \ ,$$

which is the desired bound since $|\underline{\hat{z}}_i - \underline{x}_i^*| = d(\hat{z}_i, x_i^*) = d(\hat{x}_i^{(2)}, Qx_i^*)$. By symmetry, the term $L_i^{(3)}$ is handled as $L_i^{(1)}$ and admits the same lower bound. ☐

Proof of Claim B.7. For simplicity, the notation $\underline{x}$ is dropped out in the proof of Claim B.7, and $\underline{x}$ is simply denoted by x. By definition of ϕ, we know that $(\hat{z}_i + x_i^*)/2 = (\phi(x_j^*) + x_j^*)/2$ for all $j \in I_1^{(1/2)}$, which gives the equality $\hat{z}_i + x_i^* - 2x_j^* = \phi(x_j^*) - x_j^*$. Since $0 \leq \phi(x_j^*) - x_j^* < c_l^{-1}\varepsilon_n$ for all $j \in I_1^{(1/2)-}$, we have

$$0 \leq \hat{z}_i + x_i^* - 2x_j^* \leq c_l^{-1}\varepsilon_n \ .$$

Since $c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n < 0$ for j in $I_1^{(1/2)-}$, we obtain

$$(c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)(\hat{z}_i + x_i^* - 2x_j^*) \geq (c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)c_l^{-1}\varepsilon_n \geq -c_l^{-1}\varepsilon_n^2 \ .$$

Since the number of indices in $I_1^{(1/2)-}$ is at most $1 + n_0/(2\pi)|I_1^{(1/2)-}|$ (where $|I_1^{(1/2)-}|$ is the arc length), and the length of this arc is at most $c_l^{-1}\varepsilon_n$, we conclude that

$$\sum_{j \in I_1^{(1/2)-}} (c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)(\hat{z}_i + x_i^* - 2x_j^*) \geq -\frac{n_0}{2\pi}c_l^{-2}\varepsilon_n^3 \\ = -\frac{c_e^3}{8\pi c_l^2}\sqrt{\frac{\log^3(n)}{n}} - c_l^{-1}\epsilon_n \ . \square$$

Proof of Claim B.8. Again, for convenience the notation $\underline{x}$ is dropped out here. Since all the terms in the sum are nonnegative, we can simply consider indices j in $I_1^{(1/2)+} \cap I_1^{(1/4)}$. Using $\phi(x_j^*) - x_j^* = \hat{z}_i + x_i^* - 2x_j^*$ for all $j \in I_1^{(1/2)}$ and $x_i^* = 0$, we obtain that, for $j \in I_1^{(1/4)}$, $\phi(x_j^*) - x_j^* \geq \hat{z}_i/2$. This gives

$$(c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)(\hat{z}_i + x_i^* - 2x_j^*) \geq (c_l\frac{\hat{z}_i}{2} - \varepsilon_n)\frac{\hat{z}_i}{2} \ ,$$

and, for some numerical constant $C > 0$,

$$\sum_{j \in I_1^{(1/2)+}} (c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)(\hat{z}_i + x_i^* - 2x_j^*) \geq Cn\big|I_1^{(1/2)+} \cap I_1^{(1/4)}\big|\frac{\hat{z}_i}{2}(c_l\frac{\hat{z}_i}{2} - \varepsilon_n) \ . \tag{32}$$

Since either $I_1^{(1/2)+} \subset I_1^{(1/4)}$ or $I_1^{(1/4)} \subset I_1^{(1/2)+}$ and $|I_1^{(1/4)}| = |I_1|/4$ and $|I_1^{(1/2)+}| = |I_1^{(1/2)}| - |I_1^{(1/2)-}| \geq |I_1^{(1/2)}| - c_l^{-1}\varepsilon_n = \frac{|I_1|}{2} - c_l^{-1}\varepsilon_n$, we deduce that

$$|I_1^{(1/2)+} \cap I_1^{(1/4)}| \geq \frac{|I_1|}{4} - c_l^{-1}\varepsilon_n \ .$$

Thus, we have

$$\sum_{j\in I_1^{(1/2)+}} (c_l|\phi(x_j^*) - x_j^*| - \varepsilon_n)(\hat{z}_i + x_i^* - 2x_j^*) \geq Cn\Big(\frac{|I_1|}{4} - c_l^{-1}\varepsilon_n\Big)\frac{\hat{z}_i}{2}\Big(c_l\frac{\hat{z}_i}{2} - \varepsilon_n\Big) \ .$$

Since $\hat{z}_i = |\hat{z}_i - x_i^*|$ (recall that $x_i^* = 0$), this concludes the proof. □

B.2. Proof of Proposition 3.5

Let $\mathbf{x}_S^{**}$ be a best approximation of $\mathbf{x}_S^*$ in $\mathbf{\Pi}_{n_0}$, that is, such that $d_\infty(\mathbf{x}_S^*, \mathbf{x}_S^{**}) = d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0})$. As in the proof of Lemma B.1, we restrict our attention to orthogonal transformations $Q \in \mathcal{O}$ that let $\mathcal{C}_{n_0}$ invariant. Fix i in $\overline{S}$. To prove Proposition 3.5, it suffices to establish variants of Lemmas B.2 and B.3 with

$$\widetilde{L}_i := \langle F_{i,S},\ D(Q^{-1}\hat{x}_i^{(2)}, \mathbf{x}_S^{**}) - D(x_i^*, \mathbf{x}_S^{**})\rangle, \tag{33}$$

instead of L_i. In the definition of $\widetilde{L}_i$, $\mathbf{x}_S^*$ has been replaced by $\mathbf{x}_S^{**}$.

Lemma B.9. *With probability at least $1 - 1/n^3$, we have*

$$\widetilde{L}_i \leq C_{L,e}\Big[1 + d(\hat{x}_i^{(2)}, Qx_i^*)\big(nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + n\mu_{\hat{\mathbf{x}}_S^{(1)},\mathbf{x}_S^*}(Q) + \sqrt{n\log(n)}\big)\Big] \ .$$

Lemma B.10. *For n large enough, one has*

$$\begin{aligned}\widetilde{L}_i \quad &\geq \quad C'_{l,e}nd(\hat{x}_i^{(2)}, Qx_i^*)\Big(d(\hat{x}_i^{(2)}, Qx_i^*) - \frac{\sqrt{\log(n)}}{n}\Big) - C''_{l,e}\sqrt{\frac{\log^3(n)}{n}} \\ &\quad -C_{L,e}\left\{1 + n\left[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\frac{\log(n)}{n}}\right]d(\hat{x}_i^{(2)}, Qx_i^*)\right\} \ .\end{aligned}$$

These two lemmas enforce that, with probability higher than $1 - 1/n^3$,

$$d(\hat{x}_i^{(2)}, Qx_i^*) \leq C_{l,L,e}\Big[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \mu_{\hat{\mathbf{x}}_S^{(1)},\mathbf{x}_S^*}(Q) + \sqrt{\log(n)/n}\Big] \ .$$

Indeed, assume that $d(\hat{x}_i^{(2)}, Qx_i^*) \geq C'_{l,L,e}\big(d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\log(n)/n}\big)$ where $C'_{l,L,e}$ is large enough. Then, Lemma B.10 implies that $\widetilde{L}_i \gtrsim_{c_l,c_L,c_e} nd^2(\hat{x}_i^{(2)}, Qx_i^*)$. Together with Lemma B.9, we deduce that

$$d(\hat{x}_i^{(2)}, Qx_i^*) \leq C_{l,L,e}\Big[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \mu_{\hat{\mathbf{x}}_S^{(1)},\mathbf{x}_S^*}(Q) + \sqrt{\log(n)/n}\Big] \ .$$

In any case, we conclude that

$$d(\hat{x}_i^{(2)}, Qx_i^*) \leq (C_{l,L,e} \vee C'_{l,L,e})\Big[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \mu_{\hat{\mathbf{x}}_S^{(1)},\mathbf{x}_S^*}(Q) + \sqrt{\log(n)/n}\Big] \ ,$$

with probability higher than $1 - 1/n^3$. Taking the minimum over all $Q \in \mathcal{O}$ that let $\mathcal{C}_{n_0}$ invariant and a union bound over all $i \in \overline{S}$, leads to Proposition 3.5.

B.2.1. Proof of Lemma B.9

To ease the exposition, we assume that $S = [n_0]$. Fix $i \in \overline{S}$. We start from

$$\widetilde{L}_i = \sum_{j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(x_i^*, x_j^{**})\Big) \ . \tag{34}$$

In order to come back to the setting of Lemma B.1, we replace $f(x_i^*, x_j^*)$ by $f(x_i^*, x_j^{**})$, using the bi-Lipschitz condition (5) so that

$$f(x_i^*, x_j^*) - f(x_i^*, x_j^{**}) \leq (c_L \vee c_e)\left[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\frac{\log(n)}{n}}\right] \ .$$

By triangular inequality, we have $d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(x_i^*, x_j^{**}) \leq d(Q^{-1}\hat{x}_i^{(2)}, x_i^*)$ which implies

$$\begin{aligned}\sum_{j=1}^{n_0} \Big(f(x_i^*, x_j^*) - f(x_i^*, x_j^{**})\Big)\Big(d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(x_i^*, x_j^{**})\Big)& \\ \leq (c_L \vee c_e) d(\hat{x}_i^{(2)}, Qx_i^*) r_n& \ ,\end{aligned}$$

where we define $r_n = n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}$. This leads us to

$$\widetilde{L}_i \leq (c_L \vee c_e) d(\hat{x}_i^{(2)}, Qx_i^*) r_n + \sum_{j=1}^{n_0} f(x_i^*, x_j^{**})\Big(d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(x_i^*, x_j^{**})\Big) \ .$$

Since x_j^{**} now runs over $\mathbf{\Pi}_{n_0}$, we can replace as in the proof of Lemma B.2 the sum over x_j^{**} by a sum over $Q^{-1}\hat{x}_j^{(1)}$ using a suitable permutation:

$$\begin{aligned}\widetilde{L}_i \leq\ & (c_L \vee c_e) d(\hat{x}_i^{(2)}, Qx_i^*) r_n \\ & + \sum_{j=1}^{n_0} f(x_i^*, Q^{-1}\hat{x}_j^{(1)})\Big(d(Q^{-1}\hat{x}_i^{(2)}, Q^{-1}\hat{x}_j^{(1)}) - d(x_i^*, Q^{-1}\hat{x}_j^{(1)})\Big) \ .\end{aligned}$$

The remainder of the proof follows the same lines as for Lemma B.2, except for small differences. Still, we provide some details for the sake of completeness. as in that proof we write $\hat{z}_i = Q^{-1}\hat{x}_i^{(2)}$ and $z_i^{(1)} = Q^{-1}\hat{x}_i^{(1)}$. We first apply Lemma B.4 to obtain

$$\begin{aligned}\widetilde{L}_i \leq\ & C_{L,e} d(\hat{x}_i^{(2)}, Qx_i^*)\left[n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + n\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + \sqrt{n\log(n)}\right] \\ & + \sum_{j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) \ .\end{aligned}$$

The last expression simplifies in

$$\sum_{j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(\hat{z}_i, z_j^{(1)}) - d(x_i^*, z_j^{(1)})\Big) = \langle F_{i,S},\ D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \ .$$

In Lemma B.2, we had $Qx_i^* \in \mathcal{C}_{n_0}$ so that we could use the definition of $\hat{x}_i^{(2)}$ to deduce that $\langle A_{i,S}, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle \le \langle A_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle$. Unfortunately, Qx_i^* does not necessarily belong to $\mathcal{C}_{n_0}$ anymore. To handle this minor issue, we replace x_i^* by the closest element y_i^* in $\mathcal{C}_{n_0}$. It satisfies $d(x_i^*, y_i^*) \le 2\pi/n_0$ and $Qy_i^* \in \mathcal{C}_{n_0}$. This leads us to

$$\begin{aligned}\langle A_{i,S}, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)})\rangle &\le \langle A_{i,S}, D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \\ &\le \langle A_{i,S}, D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle + \langle A_{i,S}, D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \ .\end{aligned}$$

Since $|d(Qy_i^*, x_i) - d(Qx_i^*, x_i)| \le 2\pi/n_0$ and $|f(x,y)| \le 1$, the above additional error term satisfies

$$\begin{aligned}\langle A_{i,S}, D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle &\le \langle F_{i,S}, D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \\ &\quad + \langle E_{i,S}, D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \\ &\le 2\pi + \\ &\quad C'\sqrt{\log(n)}\|D(Qy_i^*, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\|_2 \\ &\lesssim 1 \ ,\end{aligned}$$

with probability higher than $1 - 1/(2n^3)$. Putting everything together, we have shown that

$$\begin{aligned}\widetilde{L}_i \le C_{L,e} d(\hat{x}_i^{(2)}, Qx_i^*) &\left[nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + n\mu_{\hat{\mathbf{x}}_S^{(1)}, \mathbf{x}_S^*}(Q) + \sqrt{n\log(n)}\right] \\ &+ C + \langle E_{i,S},\, D(\hat{x}_i^{(2)}, \hat{\mathbf{x}}_S^{(1)}) - D(Qx_i^*, \hat{\mathbf{x}}_S^{(1)})\rangle \ ,\end{aligned}$$

with probability higher than $1 - 1/(2n^3)$. To conclude, it suffices to the rhs in the above expression. We do it exactly as in the end of the proof of Lemma B.2 except that we now consider a probability $1 - 1/(2n^3)$.

B.2.2. Proof of Lemma B.10

Fix $i \in \overline{S}$ and define $y_i^* \in \mathcal{C}_{n_0}$ as a closest point to x_i^* in $\mathcal{C}_{n_0}$. We introduce the quantity

$$L_i' = \sum_{j=1}^{n_0} f(y_i^*, x_j^{**})\big(d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(y_i^*, x_j^{**})\big) \ ,$$

which has the same properties as the L_i used in Lemma B.1, since each point involved in the expression of L_i' is an element of $\mathcal{C}_{n_0}$, and the sum runs over a vector $\mathbf{x}_S^{**}$ in $\mathbf{\Pi}_{n_0}$. This allows us to invoke Lemma B.3 –from the proof of Lemma B.1 – to get

$$L_i' \ge Cnd(\hat{x}_i^{(2)}, Qy_i^*)\left[c_l d(\hat{x}_i^{(2)}, Qy_i^*) - \varepsilon_n\right] - \frac{c_e^3}{\pi c_l^2}\sqrt{\frac{\log^3(n)}{n}} \ ,$$

for n large enough.

By definition of y_i^*, we know that $d(y_i^*, x_i^*) \leq 2\pi/n_0$. Hence, by triangular inequality, $d(\hat{x}_i^{(2)}, Qy_i^*) \geq d(\hat{x}_i^{(2)}, Qx_i^*) - 2\pi/n_0$ and we derive that

$$L_i' \geq C_{l,e}' n d(\hat{x}_i^{(2)}, Qx_i^*) \left[d(\hat{x}_i^{(2)}, Qx_i^*) - \sqrt{\frac{\log(n)}{n}} \right] - C_{l,e}'' \sqrt{\frac{\log^3(n)}{n}} \ .$$

Next, we rely on the following lemma to replace L_i' by $\widetilde{L}_i$.

Lemma B.11. *We have*

$$|\widetilde{L}_i - L_i'| \leq C_{L,e} \left[1 + \left(n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n \log(n)} \right) d(\hat{x}_i^{(2)}, Qx_i^*) \right] \ .$$

Gathering these two bounds completes the proof of Lemma B.10. ☐

Proof of Lemma B.11. From the definition of y_i^* and the triangular inequality, we have

$$\left| d(x_i^*, x_j^{**}) - d(y_i^*, x_j^{**}) \right| \leq 2\pi/n_0 \ .$$

This allows us to deduce that the quantity

$$L_i'' = \sum_{j=1}^{n_0} f(x_i^*, x_j^*) \left(d(Q^{-1}\hat{x}_i^{(2)}, x_j^{**}) - d(y_i^*, x_j^{**}) \right) \tag{35}$$

satisfies $|\widetilde{L}_i - L_i''| \leq 2\pi$. Besides, we deduce from the bi-Lipschitz condition (5) and the triangular inequality that

$$\begin{aligned} |f(x_i^*, x_j^*) - f(y_i^*, x_j^{**})| &\leq c_L \left(d(x_i^*, y_i^*) + d(x_j^*, x_j^{**}) \right) + 2\varepsilon_n \\ &\leq 2 c_L d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + 2\varepsilon_n \ . \end{aligned}$$

Then, we deduce that

$$|L_i' - L_i''| \leq 2n d(Q^{-1}\hat{x}_i^{(2)}, y_i^*) \left(c_L d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \varepsilon_n \right) \ .$$

All in all, we have

$$\begin{aligned} |\widetilde{L}_i - L_i'| &\leq |\widetilde{L}_i - L_i''| + |L_i'' - L_i'| \\ &\leq C_{L,e} \left[1 + d(Q^{-1}\hat{x}_i^{(2)}, y_i^*) \left(n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n \log(n)} \right) \right] \ , \end{aligned}$$

and the result follows. ☐

B.3. Proof of Proposition 3.6

B.3.1. Main arguments

Assume that $S = [n_0]$ for the ease of presentation. In this proof, we both interpret $\hat{\mathbf{x}}_S^{(1)}$ and $\mathbf{x}_S^*$ as vectors in $\mathcal{C}_n$ and matrices of size $2 \times n_0$. We recall that $\|.\|_q$ refers to the entry-wise l_q norm for matrices. We shall establish that the estimator $\hat{\mathbf{x}}_S^{(1)}$ is such that the matrix $\hat{\mathbf{x}}_S^{(1)T}\hat{\mathbf{x}}_S^{(1)}$ is close to $\mathbf{x}_S^{*T}\mathbf{x}_S^*$. In other words, the distances between $\hat{x}_1^{(1)}, \ldots, \hat{x}_{n_0}^{(1)}$ are close to the respective distances between the $x_1^*, \ldots, x_{n_0}^*$. Then, relying on a recent matrix perturbation result from [1], we deduce that, up to an orthogonal transformation, $\hat{\mathbf{x}}_S^{(1)}$ and $\mathbf{x}_S^*$ are close. Let us first state this perturbation result. Given any $p \times 2$ matrix M with real coefficients, we denote its transpose by M^T, and the Moore-Penrose pseudo-inverse by $M^\dagger$, and the usual operator norm by $\|M\|_{op}$. In this proof, the transformations $Q \in \mathcal{O}$ are interpreted as orthogonal matrices of size 2×2.

Proposition B.12 (Theorem 1 in [1]). *For any positive integer p and any $p\times 2$ matrices M and N, with N having full rank, let $\nu = \|MM^T - NN^T\|_2$. Then, we have*

$$\min_{Q\in\mathcal{O}} \|M - NQ\|_2 \lesssim \nu\|N^\dagger\|_{op} \ ,$$

as soon as $2\nu\|N^\dagger\|_{op}^2 \leq 1$.

Let $\mathbf{x}_S^{**} \in \mathbf{\Pi}_{n_0}$ denote a best approximation of $\mathbf{x}_S^*$ in $\mathbf{\Pi}_{n_0}$, so that $d_\infty(\mathbf{x}_S^*, \mathbf{x}_S^{**}) = d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0})$. In order to invoke the above proposition for $M = (\hat{\mathbf{x}}_S^{(1)})^T$ and $N = (\mathbf{x}_S^{**})^T$ in $\mathbb{R}^{n_0\times 2}$, we need to check that the condition $2\nu\|N^\dagger\|_{op}^2 \leq 1$ is fulfilled. First, to bound the term $\|N^\dagger\|_{op}^2$, we work out

$$N^\dagger = (N^TN)^{-1}N^T = \left(\mathbf{x}_S^{**}\mathbf{x}_S^{**T}\right)^{-1}\mathbf{x}_S^{**} = \begin{pmatrix} \sum_{i=1}^{n_0}(\mathbf{x}_{1i}^{**})^2 & 0 \\ 0 & \sum_{i=1}^{n_0}(\mathbf{x}_{2i}^{**})^2 \end{pmatrix}^{-1}\mathbf{x}_S^{**}$$
$$= \frac{2}{n_0}\mathbf{x}_S^{**} \ ,$$

since $\sum_{i=1}^{n_0}(\mathbf{x}_{1i}^{**})^2 = \sum_{i=1}^{n_0}(\mathbf{x}_{2i}^{**})^2 = n_0/2$ – see e.g. (76) for a proof. As a consequence,

$$\|N^\dagger\|_{op}^2 \leq \|N^\dagger\|_2^2 = \frac{4}{n_0} \ .$$

The following lemma bounds $\nu = \|\hat{\mathbf{x}}_S^{(1)T}\hat{\mathbf{x}}_S^{(1)} - \mathbf{x}^{**T}\mathbf{x}_S^{**}\|_2$.

Lemma B.13. *With probability at least $1 - 1/n^2$, we have*

$$\nu \leq C_{l,L,e}\left(nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}\right) \ .$$

Hence, with probability higher than $1 - 1/n^2$, we obtain

$$2\nu\|N^\dagger\|_{op}^2 \leq C'_{l,L,e}\left(d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\frac{\log(n)}{n}}\right) \ .$$

If $d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\log(n)/n} \geq 1/C'_{l,L,e}$, the conclusion of Proposition 3.6 obviously holds since $\min_{Q\in\mathcal{O}} d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}_S^*) \leq n_0 \leq n$. Hence, it suffices to consider the case where $d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + \sqrt{\log(n)/n} \leq 1/C'_{l,L,e}$ so that the condition of Proposition B.12 is fulfilled. This implies

$$\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{(1)T} - \mathbf{x}^{**T}Q\|_2 \lesssim C\left(\sqrt{n}d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + \sqrt{\log(n)}\right) \quad ,$$

and so

$$\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{(1)} - Q\mathbf{x}^{**}_S\|_1 \lesssim \sqrt{2}C\left(nd_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}\right) \quad .$$

Since the distances in $\mathbb{R}^2$ and $\mathcal{C}$ are equivalent, we have $\|Q\mathbf{x}^{**}_S - Q\mathbf{x}^*_S\|_1 \lesssim d_1(Q\mathbf{x}^{**}_S, Q\mathbf{x}^*_S) = d_1(\mathbf{x}^{**}_S, \mathbf{x}^*_S)$. Then, using the definition of $d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0})$, we get $\|Q\mathbf{x}^{**}_S - Q\mathbf{x}^*_S\|_1 \lesssim nd_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0})$. Together with the triangular inequality, this leads us to

$$\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{(1)} - Q\mathbf{x}^*_S\|_1 \leq C_{l,L,e}\left[nd_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}\right] \quad .$$

Using again the equivalence between the distances, that is $d(x,y) \lesssim \|x-y\|_1$ for all $x, y \in \mathcal{C}$, we conclude that

$$\min_{Q\in\mathcal{O}} d_1(\hat{\mathbf{x}}_S^{(1)}, Q\mathbf{x}^*_S) \leq C \min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{(1)} - Q\mathbf{x}^*\|_1 \leq C_{l,L,e}\left[nd_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}\right] \quad ,$$

and the proof of Proposition 3.6 is complete.

B.3.2. Proof of Lemma B.13

Both $\hat{\mathbf{x}}_S^{(1)}$ and $\mathbf{x}^{**}_S$ are elements of $\mathbf{\Pi}_{n_0} \subset \mathbb{R}^{2\times n_0}$, hence they both satisfy $\hat{\mathbf{x}}_S^{(1)}1 = 0$ and $\mathbf{x}^{**}_S 1 = 0$ where 1 denotes the vector of ones. Indeed, since $\sum_{k=0}^{n_0-1} e^{\iota 2\pi k/n_0} = 0$, we have $\sum_{k=0}^{n_0-1}\cos(2\pi k/n_0) = 0$ and $\sum_{k=0}^{n_0-1}\sin(2\pi k/n_0) = 0$. We can then invoke the next lemma to bound $\nu = \|\hat{\mathbf{x}}_S^{(1)T}\hat{\mathbf{x}}_S^{(1)} - \mathbf{x}^{**T}\mathbf{x}^{**}_S\|_2$.

Lemma B.14. *For any $Z = (z_1, \dots, z_{n_0})$ and $Z' = (z'_1, \dots, z'_{n_0})$ in $\mathbb{R}^{2\times n_0}$ with $Z1 = Z'1 = 0$, let $D = (D_{ij})$ and $D' = (D'_{ij})$ be their (squared) distance matrices, that is $D_{ij} = \|z_i - z_j\|_2^2$ and $D'_{ij} = \|z'_i - z'_j\|_2^2$ for all $i, j \in [n_0]$. Then we have*

$$\|Z^TZ - Z'^TZ'\|_2 \leq \|D - D'\|_2 \quad .$$

For $Z = \mathbf{x}^{**}_S$ and $Z' = \hat{\mathbf{x}}_S^{(1)}$, and accordingly $D_{ij} = \|x^{**}_i - x^{**}_j\|_2^2$ and $D'_{ij} = \|\hat{x}_i^{(1)} - \hat{x}_j^{(1)}\|_2^2$, it follows from Lemma B.14 that $\nu \leq \|D - D'\|_2$. Since all square distances D_{ij} and D'_{ij} are at most equal to 4, we get

$$\nu \leq 4\|\sqrt{D} - \sqrt{D'}\|_2 \quad ,$$

where $\sqrt{D}$ and $\sqrt{D'}$ denote the matrices of coefficients $\sqrt{D_{ij}} = \|x^{**}_i - x^{**}_j\|_2$ and $\sqrt{D'_{ij}} = \|\hat{x}_i^{(1)} - \hat{x}_j^{(1)}\|_2$.

For any $x, y \in \mathcal{C}$, elementary geometry gives $\|x-y\|_2 = 2\sin(d(x,y)/2)$. Since the sinus function is 1-Lipschitz, we have

$$\big|\,\|x-y\|_2 - \|x'-y'\|_2\big| \leq |d(x,y) - d(x',y')| \ ,$$

for any $x, y, x', y' \in \mathcal{C}$. Hence, we deduce that $\nu \leq 4\|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2$, where $D(\mathbf{x}_S^{**})$ and $D(\hat{\mathbf{x}}_S^{(1)})$ respectively denote the matrices of coefficients $d(x_i^{**}, x_j^{**})$ and $d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)})$. As a consequence, we mainly have to control with high probability $\|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2$.

Lemma B.15. *With probability at least* $1 - 1/n^2$, *we have*

$$\|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2 \leq C'_{l,L,e}\left[nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}\right] \ .$$

Hence $\nu \leq C_{l,L,e}[nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{n\log(n)}]$ and the proof of Lemma B.13 is complete. □

Proof of Lemma B.14. Let $H = I - J/n_0$, where I is the identity and J the matrix of ones. Since $Z1 = 0$, we have $ZH = Z$, so that

$$Z^TZ = HZ^TZH = -\frac{1}{2}HDH,$$

since D is the matrix of distances associated with Z. Then we have

$$\|Z^TZ - Z'^TZ'\|_2 = \frac{1}{2}\|H(D-D')H\|_2 \leq \frac{1}{2}\|D-D'\|_2 \ ,$$

where the last inequality derives from the general relation $\|AB\|_2 \leq \|A\|_{op}\|B\|_2$ for any matrices A, B, and the fact that $\|H\|_{op} = 1$ – because H is an orthogonal projection. Lemma B.14 is proved. □

Proof of Lemma B.15. First, we come back to the definition of the estimator $\hat{\mathbf{x}}_S^{(1)}$ defined in (18). We have $\langle A_{S,S}, D(\hat{\mathbf{x}}_S^{(1)})\rangle \leq \langle A_{S,S}, D(\mathbf{x}_S^{**})\rangle$, which implies that

$$\langle F_{S,S}, D(\hat{\mathbf{x}}_S^{(1)}) - D(\mathbf{x}_S^{**})\rangle \leq \langle E_{S,S}, D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\rangle.$$

As in the last lines of the proof of Lemma B.2, we bound the term $\langle E_{S,S}, D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\rangle$ by a union bound over all possible vectors $\hat{\mathbf{x}}_S^{(1)}$. Hence, we get

$$\begin{aligned}\langle F_{S,S}, D(\hat{\mathbf{x}}_S^{(1)}) - D(\mathbf{x}_S^{**})\rangle &\leq \langle E_{S,S}, D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\rangle \\ &\lesssim \sqrt{n\log(n)}\|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2 \ , \end{aligned} \tag{36}$$

with probability at least 1–1/n^2. Conversely, we shall lower bound $\langle F_{S,S}, D(\hat{\mathbf{x}}_S^{(1)}) - D(\mathbf{x}_S^{**})\rangle$.

$$\langle F_{S,S}, D(\hat{\mathbf{x}}_S^{(1)}) - D(\mathbf{x}_S^{**})\rangle = \sum_{i,j=1}^{n_0} f(x_i^*, x_j^*)\Big(d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**})\Big) \ .$$

Using the bi-Lipschitz property of the function f, we deduce that

$$\begin{aligned}\left|f(x_i^*,x_j^*)-f(x_i^{**},x_j^{**})\right| &\leq c_L\left(d(x_i^*,x_i^{**})+d(x_j^*,x_j^{**})\right)+2c_e\sqrt{\log(n)/n}\\ &\leq 2c_L d_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+2c_e\sqrt{\log(n)/n}\ ,\end{aligned}$$

by definition of $\mathbf{x}^{**}$. Then, we get

$$\begin{aligned}&\textstyle\sum_{i,j=1}^{n_0}\left|f(x_i^*,x_j^*)-f(x_i^{**},x_j^{**})\right|\left|d(\hat{x}_i^{(1)},\hat{x}_j^{(1)})-d(x_i^{**},x_j^{**})\right|\\ &\quad\leq C_{L,e}\left(d_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+\sqrt{\tfrac{\log(n)}{n}}\right)\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_1\\ &\quad\leq C_{L,e}n_0\left(d_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+\sqrt{\tfrac{\log(n)}{n}}\right)\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2\quad ,\end{aligned}$$

where we applied Cauchy-Schwarz inequality on $\mathbb{R}^{n_0\times n_0}$. As a consequence,

$$\begin{aligned}\langle F_{S,S},D(\hat{\mathbf{x}}_S^{(1)})-D(\mathbf{x}_S^{**})\rangle &\geq \sum_{i,j=1}^{n_0} f(x_i^{**},x_j^{**})\Big(d(\hat{x}_i^{(1)},\hat{x}_j^{(1)})-d(x_i^{**},x_j^{**})\Big)\\ &\quad -C_{L,e}n\left(d_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+\sqrt{\frac{\log(n)}{n}}\right)\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2\ . \end{aligned}\tag{37}$$

The following result bounds $\sum_{i,j=1}^{n_0} f(x_i^{**},x_j^{**})[d(\hat{x}_i^{(1)},\hat{x}_j^{(1)})-d(x_i^{**},x_j^{**})]$ in terms of the Frobenius norm $\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2^2$. This is a key step in our proof. Had the slack constant c_e been equal to zero, the following result would have been a consequence of Lemma A.1. Here the proof is slightly more involved and is provided below.

Lemma B.16. *We have*

$$\begin{aligned}\sum_{i,j=1}^{n_0} f(x_i^{**},x_j^{**})\Big(d(\hat{x}_i^{(1)},\hat{x}_j^{(1)})-d(x_i^{**},x_j^{**})\Big) &\geq \frac{c_l}{2}\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2^2\\ &\quad -c_e\sqrt{n\log(n)}\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2.\end{aligned}$$

We conclude from (36) and the above lemma that

$$\begin{aligned}C\sqrt{n\log(n)}\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2 &\geq \frac{c_l}{2}\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2^2\\ &\quad -C_{L,e}\left(nd_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+\sqrt{n\log(n)}\right)\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2\end{aligned}$$

which in turn implies that

$$\|D(\mathbf{x}_S^{**})-D(\hat{\mathbf{x}}_S^{(1)})\|_2\leq C'_{l,L,e}\left[nd_\infty(\mathbf{x}_S^*,\mathbf{\Pi}_{n_0})+\sqrt{n\log(n)}\right]\ .$$

Lemma B.15 is proved. ∎

Proof of Lemma B.16. To alleviate the notation, we introduce

$\gamma_i = \sum_{j=1}^{n_0} f(x_i^{**}, x_j^{**})[d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**})]$ so that we aim at establishing a lower bound for each γ_i and in turn for $\gamma = \sum_{i=1}^{n_0} \gamma_i$. To simplify the arguments, we only consider the case where n_0 is odd, the case of n_0 even being almost similar.

Both $\mathbf{x}_S^{**}$ and $\hat{\mathbf{x}}_S^{(1)}$ belongs to $\mathbf{\Pi}_{n_0}$ and we shall heavily rely on the symmetries of $\mathbf{\Pi}_{n_0}$. Assume without loss of generality that $i = 1$ and $x_j^{**} = e^{\iota 2\pi(j-1)/n_0}$ for all $j = 1, \ldots, n_0$. Then, $d(x_i^{**}, x_j^{**}) = \frac{2\pi}{n_0}[|j-1| \wedge |n_0 - j + 1|]$. Since $\hat{\mathbf{x}}_S^{(1)}$ also belongs to $\mathbf{\Pi}_{n_0}$, there exists a permutation σ of $[n_0]$ such that $\sigma(1) = 1$ and $d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) = \frac{2\pi}{n_0}[|\sigma(j) - 1| \wedge |n_0 + 1 - \sigma(j)|]$. Recall that we consider the case where n_0 is odd. Besides, we can focus on n_0 larger than 3 since Lemma B.16 is trivial for $n_0 = 1$. Thus, there exists a surjective map $\overline{\sigma} : [n_0 - 1] \mapsto [\lfloor n_0/2 \rfloor]$ such that $|\overline{\sigma}^{-1}(\{z\})| = 2$ for any $z \in [\lfloor n_0/2 \rfloor]$ and $d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) = \frac{2\pi}{n_0}\overline{\sigma}(j-1)$ for any $j = 2, \ldots n_0$. Finally, we write $\psi_j = f(1, e^{\iota 2\pi j/n_0})$ and $\psi'_j = f(1, e^{-\iota 2\pi j/n_0})$ for $j = 1, \ldots \lfloor n_0/2 \rfloor$. Equipped with this new notation, we arrive at

$$\gamma_i = \frac{2\pi}{n_0} \sum_{j=1}^{\lfloor n_0/2 \rfloor} \psi_j[\overline{\sigma}(j) - j] + \psi'_j[\overline{\sigma}(n_0 - j) - j] \ .$$

Finally, we denote $a_j = \overline{\sigma}(j) - j$ and $a'_j = \overline{\sigma}(n_0 - j) - j$ for $j = 1, \ldots, \lfloor n_0/2 \rfloor$. Obviously, we have $\sum_{j=1}^{\lfloor n_0/2 \rfloor} a_j + a'_j = 0$. More generally, one easily checks that, for any positive integer $s \leq \lfloor n_0/2 \rfloor$, the sum $\sum_{j=1}^{s}(a_j + a'_j)$ is nonnegative. Starting from

$$\gamma_i = \frac{2\pi}{n_0} \sum_{j=1}^{\lfloor n_0/2 \rfloor} \psi_j a_j + \psi'_j a'_j$$

we partition the indices according to the signs of a_j and a'_j. Define $A_+ = \{j \in [\lfloor n_0/2 \rfloor] : a_j \geq 0\}$, $A_- = \{j \in [\lfloor n_0/2 \rfloor] : a_j < 0\}$, $A'_+ = \{j \in [\lfloor n_0/2 \rfloor] : a'_j \geq 0\}$, and $A'_- = \{j \in [\lfloor n_0/2 \rfloor] : a'_j < 0\}$. Intuitively, we want to group indices j such that $a_j > 0$ with indices k such that $a_k < 0$. This can be done by recursion. First, consider the smallest index $k \in A_- \cup A'_-$. By symmetry, suppose that $a_k < 0$. Since $\sum_{j=1}^{k}(a_j + a'_j) \geq 0$, this implies that $\sum_{j=1}^{k} \mathbf{1}_{j \in A_+} a_j + \mathbf{1}_{j \in A'_+} a'_j \geq |a_k| + \mathbf{1}_{k \in A'_-} |a'_k|$. Hence, it is possible to build nonnegative numbers $b_{j,k,1} \leq a_j$ for $j \in A_+ \cap [k]$ and $b'_{j,k,1} \leq a'_j$ for $j \in A'_+ \cap [k]$ such that $\sum_{j=1}^{k} \mathbf{1}_{j \in A_+} b_{j,k,1} + \mathbf{1}_{j \in A'_+} b'_{j,k,1} = |a_k|$. Iterating the construction we obtain the following decomposition

$$\begin{aligned}
\frac{n_0}{2\pi}\gamma_i \quad = \quad & \sum_{j \in A_+} \left(\sum_{k \in A_-} (\psi_j - \psi_k) b_{j,k,1} + \sum_{k \in A'_-} (\psi_j - \psi'_k) b_{j,k,2} \right) \\
& + \sum_{j \in A'_+} \left(\sum_{k \in A_-} (\psi'_j - \psi_k) b'_{j,k,1} + \sum_{k \in A'_-} (\psi'_j - \psi'_k) b'_{j,k,2} \right) \ ,
\end{aligned}$$

where all $b_{j,k,t}$'s are nonnegative, $b_{j,k,t} = 0$ for $k < j$, and

$$\begin{cases} \sum_{k\in A_-} b_{j,k,1} + \sum_{k\in A'_-} b_{j,k,2} = a_j \text{ for } j \in A_+ \ ; \\ \sum_{k\in A_-} b'_{j,k,1} + \sum_{k\in A'_-} b'_{j,k,2} = a'_j \text{ for } j \in A'_+ \ ; \\ \sum_{j\in A_+} b_{j,k,1} + \sum_{j\in A'_+} b'_{j,k,1} = -a_k \text{ for } k \in A_- \ ; \\ \sum_{j\in A_+} b_{j,k,2} + \sum_{j\in A'_+} b'_{j,k,2} = -a'_k \text{ for } k \in A'_- \ . \end{cases}$$

In the above decomposition all the terms $b_{j,k,1}$, $b_{j,k,2}$, $b'_{j,k,1}$, and $b'_{j,k,2}$ are nonnegative. Besides, they are positive only when $k \geq j$, so that we can use the bi-Lipschitz condition (6)

$$(\psi_j - \psi'_k) = f(1, e^{\iota 2\pi(j-1)/n_0}) - f(1, e^{-\iota 2\pi(k-1)/n_0}) \geq c_l \frac{2\pi(k-j)}{n_0} - c_e\sqrt{\frac{\log(n)}{n}} \ .$$

We obtain similarly the same lower bound for $\psi_j - \psi_k$, $\psi'_j - \psi_k$, and $\psi'_j - \psi'_k$. Coming back to the expression γ_j and the definition of the $b_{i,j,t}$ with $t = 1, 2$ yields

$$\begin{aligned} \frac{n_0}{2\pi}\gamma_i \quad &\geq \quad c_l \frac{2\pi}{n_0} \sum_{j=1}^{\lfloor n_0/2\rfloor} -j\left[a_j + a'_j\right] - c_e\sqrt{\frac{\log(n)}{n}} \sum_{j=1}^{\lfloor n_0/2\rfloor} |a_j| + |a'_j| \\ &\geq \quad -c_l \frac{2\pi}{n_0} \sum_{j=1}^{\lfloor n_0/2\rfloor} \left[j(\overline{\sigma}(j) - j) + j(\overline{\sigma}(n_0 - j) - j)\right] \\ &\qquad -c_e\sqrt{\frac{\log(n)}{n}} \sum_{j=1}^{\lfloor n_0/2\rfloor} |\overline{\sigma}(j) - j| + |\overline{\sigma}(n_0 - j) - j| \ . \end{aligned}$$

Let us work out these two expressions in the rhs. By symmetry and definition of $\overline{\sigma}$ and σ we get

$$\begin{aligned} &\sum_{j=1}^{\lfloor n_0/2\rfloor} -j(\overline{\sigma}(j) - j) - j(\overline{\sigma}(n_0 - j) - j) \\ &\qquad = \frac{1}{2} \sum_{j=1}^{\lfloor n_0/2\rfloor} (\overline{\sigma}(j) - j)^2 + (\overline{\sigma}(n_0 - j) - j)^2 \\ &\qquad = \frac{n_0^2}{8\pi^2} \sum_{j=1}^{n_0} \left[d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**})\right]^2 \ . \end{aligned}$$

Similarly, we get

$$\sum_{j=1}^{\lfloor n_0/2\rfloor} |\sigma(j) - j| + |\overline{\sigma}(n_0 - j) - j| = \frac{n_0}{2\pi} \sum_{j=1}^{n_0} |d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**})| \ .$$

Putting everything together yields

$$\gamma_i \geq \frac{c_l}{2} \sum_{j=1}^{n_0} \left[d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**}) \right]^2 - c_e \sqrt{\frac{\log(n)}{n}} \sum_{j=1}^{n_0} |d(\hat{x}_i^{(1)}, \hat{x}_j^{(1)}) - d(x_i^{**}, x_j^{**})| \ ,$$

which in turn allows us to conclude

$$\begin{aligned} \gamma &\geq \frac{c_l}{2} \|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2^2 - c_e \sqrt{\log(n)/n} \|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_1 \\ &\geq \frac{c_l}{2} \|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2^2 - c_e \sqrt{n \log(n)} \|D(\mathbf{x}_S^{**}) - D(\hat{\mathbf{x}}_S^{(1)})\|_2 \ . \end{aligned}$$

Lemma B.16 is proved. □

B.4. Proof of Theorem 3.1 and 3.2

In this section, we prove Theorem 3.2. Theorem 3.1 then follows directly from this result.

B.4.1. Main arguments

Recall that $n = 4n_0$. For $n_0 \leq 3$, the bound of Theorem 3.2 is trivially true. Assume that $n_0 \geq 4$ in the following. In Step 1 of the main procedure, it follows from Propositions 3.5 and 3.6 that the output $\hat{\mathbf{x}}_{\overline{S}}^{(2)}$ satisfies the following uniform bound

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}_{\overline{S}}^{(2)}, Q\mathbf{x}_{\overline{S}}^*) \leq C_{l,L,e} \left[d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + \sqrt{\frac{\log(n)}{n}} \right] \ , \tag{38}$$

with probability higher than $1 - 2/n^2$. Similarly, for the output $\hat{\mathbf{x}}_{\overline{S'}}^{(2')}$ in Step 2, we have

$$\min_{Q \in \mathcal{O}} d_\infty(\hat{\mathbf{x}}_{\overline{S'}}^{(2')}, Q\mathbf{x}_{\overline{S'}}^*) \leq C_{l,L,e} \left[d_\infty(\mathbf{x}_{S'}^*, \mathbf{\Pi}_{n_0}) + \sqrt{\frac{\log(n)}{n}} \right] \ , \tag{39}$$

with probability higher than $1 - 2/n^2$. In (38) and (39), we shall prove that the bias terms $d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0})$ and $d_\infty(\mathbf{x}_{S'}^*, \mathbf{\Pi}_{n_0})$ are of the same order as $d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n)$ up to an additional error of the order of $\sqrt{\log(n)/n}$ – see Lemma B.17 below.

Lemma B.17. *Assume that $n_0 \geq 4$ and fix $\mathbf{x}^* \in \mathcal{C}^n$. There exists an event of probability higher than $1 - 1/n^2$ such that*

$$d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) \leq C \left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}} \right] \ ;$$

$$d_\infty(\mathbf{x}^*_{S'}, \mathbf{\Pi}_{n_0}) \leq C\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ ;$$

$$d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \mathbf{\Pi}_{2n_0}) \leq C\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ .$$

Thus, by a union bound, the following inequalities hold together with probability at least $1 - 5/n^2$:

$$\exists\, Q_1 \in \mathcal{O}: \qquad d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}}, Q_1\mathbf{x}^*_{\overline{S}}) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ , \tag{40}$$

$$\exists\, Q_2 \in \mathcal{O}: \qquad d_\infty(\hat{\mathbf{x}}^{(2')}_{\overline{S}'}, Q_2\mathbf{x}^*_{\overline{S}'}) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ . \tag{41}$$

Since the final estimator $\hat{\mathbf{x}} := (\hat{\mathbf{x}}^{(2)}_{\overline{S}}, \widehat{Q}\hat{\mathbf{x}}^{(2')}_{S})$ satisfies $\hat{\mathbf{x}}_{\overline{S}} = \hat{\mathbf{x}}^{(2)}_{\overline{S}}$, we deduce from (40) that

$$d_\infty(\hat{\mathbf{x}}_{\overline{S}}, Q_1\mathbf{x}^*_{\overline{S}}) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ .$$

To prove Theorem 3.2, it suffices to show the counterpart of this bound on S:

$$d_\infty(\hat{\mathbf{x}}_{S}, Q_1\mathbf{x}^*_{S}) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ . \tag{42}$$

By the triangle inequality, we have

$$\begin{aligned} d_\infty(\hat{\mathbf{x}}_S, Q_1\mathbf{x}^*_S) &\leq d_\infty(\hat{\mathbf{x}}_S, \widehat{Q}Q_2\mathbf{x}^*_S) + d_\infty(Q_1\mathbf{x}^*_S, \widehat{Q}Q_2\mathbf{x}^*_S) \\ &= d_\infty(\hat{\mathbf{x}}^{(2')}_S, Q_2\mathbf{x}^*_S) + \max_{y\in\mathcal{C}} d(Q_1y, \widehat{Q}Q_2y) \ , \end{aligned} \tag{43}$$

since $\hat{\mathbf{x}}_S = \widehat{Q}\hat{\mathbf{x}}^{(2')}_S$ by definition of $\hat{\mathbf{x}}$. By (41) and since $S \subset \overline{S}'$, we have

$$d_\infty(\hat{\mathbf{x}}_S, \widehat{Q}Q_2\mathbf{x}^*_S) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ .$$

In view of (42) and (43), it remains to prove that

$$\max_{y\in\mathcal{C}} d(Q_1y, \widehat{Q}Q_2y) \leq C_{l,L,e}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ . \tag{44}$$

Before consider this maximum, we control the quantity $d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'})$ that will turn out to be instrumental. By the triangular inequality,

$$d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) \quad \leq \quad d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S}'}) + d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S}'}, \widehat{Q}\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S}'})$$

$$+ \quad d_\infty(\widehat{Q}\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) \ .$$

By definition of $\widehat{Q}$, the second term of the right hand-side is bounded by $d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S}'}, Q_1Q_2^{-1}\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S}'})$, which, in turn, is bounded as follows

$$d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S}'}, Q_1Q_2^{-1}\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S}'}) \leq d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}\cap\overline{S}'}, Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) + d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, Q_1Q_2^{-1}\hat{\mathbf{x}}^{(2')}_{\overline{S}\cap\overline{S}'}) \ .$$

Together with (40) and (41), this leads us to

$$\begin{aligned} d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) &\leq 2d_\infty(\hat{\mathbf{x}}^{(2)}_{\overline{S}}, Q_1\mathbf{x}^*_{\overline{S}}) + 2d_\infty(\hat{\mathbf{x}}^{(2')}_{\overline{S}'}, Q_2\mathbf{x}^*_{\overline{S}'}) \\ &\leq C_{lLe}\left[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right] \ . \qquad (45) \end{aligned}$$

Let us now come back to proving (44). Since the symmetric group on the plane is only made of rotations and reflections, we consider two cases.

Case 1: $Q_1^{-1}\widehat{Q}Q_2$ is a rotation. Then, $d(Q_1y, \widehat{Q}Q_2y)$ does not depend on y. In particular, $\max_{y\in\mathcal{C}} d(Q_1y, \widehat{Q}Q_2y) = d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'})$ and (44) is a consequence of (45).

Case 2: $Q_1^{-1}\widehat{Q}Q_2$ is a reflection. Then, $\max_{y\in\mathcal{C}} d(Q_1y, \widehat{Q}Q_2y) = \pi$.

If $d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, Q_1^{-1}\widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) \geq \pi/4$, then

$$\max_{y\in\mathcal{C}} d(Q_1y, \widehat{Q}Q_2y) \leq 4d_\infty(Q_1\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) \ ,$$

and (44) is again a consequence of (45). If $d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, Q_1^{-1}\widehat{Q}Q_2\mathbf{x}^*_{\overline{S}\cap\overline{S}'}) \leq \pi/4$, this implies that the points in $\mathbf{x}^*_{\overline{S}\cap\overline{S}'}$ belong to two arcs of length $\pi/4$ that are (individually) symmetric around the axis of the reflection $Q_1^{-1}\widehat{Q}Q_2$. It follows that $d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \mathbf{\Pi}_{2n_0}) \geq \pi/8$ as soon as $2n_0 \geq 4$, that is $n \geq 8$. Indeed, if $d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \mathbf{\Pi}_{2n_0}) < \pi/8$ and $2n_0 \geq 4$, this would imply that, any point on $\mathcal{C}$ is at distance less than $3\pi/8$ from $\mathbf{x}^*_{\overline{S}\cap\overline{S}'}$ which is impossible because those points in $\mathbf{x}^*_{\overline{S}\cap\overline{S}'}$ belong to these two arcs of length $\pi/4$. Since Lemma B.17 ensures that $d_\infty(\mathbf{x}^*_{\overline{S}\cap\overline{S}'}, \mathbf{\Pi}_{2n_0})$ is of the same order as $d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n)$, this implies that the latter is of the order of a constant and (44) is obviously valid.

B.4.2. Proof of Lemma B.17

We claim that it suffices to restrict our attention to the case where $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ are n distinct points. Indeed, for general points $x_1^*, \dots, x_n^*$ in $\mathcal{C}$, there exist points $y_1, \dots, y_n$ that are all distinct and satisfy $d(y_j, x_j^*) \leq 1/n$ for all $j \in [n]$. Replacing $x_1^*, \dots, x_n^*$ by $y_1, \dots, y_n$ in the statement of Lemma B.17 only entails an additional term $1/n$ which is negligible compared to the term $\sqrt{\log(n)/n}$.

For any $k \in [n]$ and any vector $\mathbf{x} \in \mathcal{C}^k$, we introduce a new quantity that is equivalent to $d_\infty(\mathbf{x}, \mathbf{\Pi}_k)$, but more easy to handle. For any interval $I \subset \mathbb{R}/(2\pi)$, we write $N_I(\mathbf{x})$ the number of coordinates of $\mathbf{x}$ that lie in the interval I, i.e. the number of $i \in [k]$ such that $\underline{x}_i \in I$. We then define the quantity $V_I(\mathbf{x})$ as

$$V_I(\mathbf{x}) = N_I(\mathbf{x}) - k\frac{|I|}{2\pi} \ . \tag{46}$$

Remark that, for a uniform k-sample of $\mathcal{C}$, the fraction $k|I|/(2\pi)$ would be the expected number of points in I. The next lemma shows that the supremum $\sup_I V_I(\mathbf{x})$ is equivalent to $k\, d_\infty(\mathbf{x}, \mathbf{\Pi}_k)$. We note $\mathcal{I}$ the set of all closed intervals $I \subset \mathbb{R}/(2\pi)$.

Lemma B.18. *For any integer $k \in [n]$ and any vector $\mathbf{x} = (x_1, \dots, x_k)$ of k distinct points of $\mathcal{C}$, we have*

$$\sup_{I \in \mathcal{I}} |V_I(\mathbf{x})| - 4 \leq \frac{k}{\pi} d_\infty(\mathbf{x}, \mathbf{\Pi}_k) \leq 2 \sup_{I \in \mathcal{I}} |V_I(\mathbf{x})| + 4 \ .$$

Thus, to prove Lemma B.17, it is enough to show that, for $T = S$, S', and $\overline{S} \cap \overline{S}'$, one has

$$\mathbb{P}\left[\sup_{I \in \mathcal{I}} \left| |V_I(\mathbf{x}^*_T)| - \frac{|T|}{n} |V_I(\mathbf{x}^*)| \right| > C\sqrt{n \log(n)}\right] \leq \frac{1}{n^3} \ . \tag{47}$$

The next Lemma states a uniform concentration bound for V_I.

Lemma B.19. *Consider any integer $n > 4$ and any integer $k < n$. Fix any $\mathbf{x} \in \mathcal{C}^n$. Sampling uniformly at random k coordinates of $\mathbf{x}$ without replacement, we write $\mathbf{x}^{(k)} \in \mathcal{C}^k$ the resulting vector. Then, with probability higher than $1 - 1/n^3$, one has*

$$\sup_{I \in \mathcal{I}} \left| \left| V_I(\mathbf{x}^{(k)}) \right| - \frac{k}{n} |V_I(\mathbf{x})| \right| \leq 6\sqrt{n \log(n)} \ . \tag{48}$$

Since the marginal distributions of S, S', and $\overline{S} \cap \overline{S}'$ are uniform, we can apply Lemma B.19 to $\mathbf{x}^*_S$, $\mathbf{x}^*_{S'}$, and $\mathbf{x}^*_{\overline{S} \cap \overline{S}'}$ and the conclusion of the Lemma holds with probability higher than $1 - 3/n^3$, which is higher than $1 - 1/n^2$. ☐

Proof of Lemma B.19. We start with a fixed interval $I \in \mathcal{I}$. Since $N_I(\mathbf{x}^{(k)})$ is a hypergeometric random variable with parameters $(k, \frac{N_I(\mathbf{x})}{n}, n)$, we can invoke Hoeffding inequality (82) for hypergeometric distributions and get

$$\mathbb{P}\left(\left| N_I(\mathbf{x}^{(k)}) - k\frac{N_I(\mathbf{x})}{n} \right| \geq \sqrt{\frac{7k \log(n)}{2}} \right) \leq \frac{2}{n^7} \leq \frac{1}{n^6} \ . \tag{49}$$

We combine (49) with

$$N_I(\mathbf{x}^{(k)}) - k\frac{|I|}{2\pi} = \left(N_I(\mathbf{x}^{(k)}) - k\frac{N_I(\mathbf{x})}{n} \right) + \frac{k}{n}\left(N_I(\mathbf{x}) - n\frac{|I|}{2\pi} \right) \ ,$$

to conclude that

$$\mathbb{P}\left[\left| |V_I(\mathbf{x}^{(k)})| - \frac{k}{n} |V_I(\mathbf{x})| \right| \leq \sqrt{\frac{7n \log(n)}{2}} \right] \leq \frac{1}{n^6} \ . \tag{50}$$

In order to extend (50) to all intervals $I \in \mathcal{I}$, we use an ε-net approach with a subcollection $\mathcal{I}_n(\mathbf{x})$ of $\mathcal{I}$. Let $\mathcal{I}_n(\mathbf{x})$ be the collection of all intervals $I_n = [a_n, b_n]$ where $a_n, b_n \in \{\underline{x}_1 \ldots, \underline{x}_n\} \cup \mathcal{C}_n$, i.e., a_n, b_n are either coordinates of $\mathbf{x}$ or elements of the n-regular grid $\{2\pi i/n;\ i \in [n]\}$. We then apply (50) together with a union bound over all intervals $I \in \mathcal{I}_n(\mathbf{x})$. Since $|\mathcal{I}_n(\mathbf{x})| \leq (2n)^2 \leq n^3$, we obtain

$$\sup_{I \in \mathcal{I}_n(\mathbf{x})} \left| \left| V_I(\mathbf{x}^{(k)}) \right| - \frac{k}{n} \left| V_I(\mathbf{x}) \right| \right| \leq \sqrt{\frac{7n \log(n)}{2}} \ , \tag{51}$$

with probability higher than $1 - 1/n^3$.

To obtain (51) for all $I \in \mathcal{I}$, we observe that, for any $I \in \mathcal{I}$, there exists $I_n \in \mathcal{I}_n(\mathbf{x})$ such that

$$I = I^{(l)} \cup I_n \cup I^{(r)}, \tag{52}$$

where $I^{(l)}$ and $I^{(r)}$ are two closed intervals of $I \setminus I_n$ whose lengths are smaller than $2\pi/n$ and that satisfy $N_{I^{(l)}}(\mathbf{x}) = N_{I^{(r)}}(\mathbf{x}) = 0$. In particular, we have $N_{I^{(l)}}(\mathbf{x}^{(r)}) = N_{I^{(l)}}(\mathbf{x}^{(k)}) = 0$. We then deduce that

$$\begin{aligned} V_I(\mathbf{x}^{(k)}) &= V_{I^{(l)}}(\mathbf{x}^{(k)}) + V_{I_n}(\mathbf{x}^{(k)}) + V_{I^{(r)}}(\mathbf{x}^{(k)}) \\ &= -k\frac{|I^{(l)}|}{2\pi} + V_{I_n}(\mathbf{x}^{(k)}) - k\frac{|I^{(r)}|}{2\pi} \ . \end{aligned}$$

Since the same decomposition holds for $V_I(\mathbf{x})$, we get

$$\left| |V_I(\mathbf{x}^{(k)})| - \frac{k}{n} |V_I(\mathbf{x})| \right| \leq 4 + \left| |V_{I_n}(\mathbf{x}^{(k)})| - \frac{k}{n} |V_{I_n}(\mathbf{x})| \right|$$

Together with (51), we obtain

$$\sup_{I \in \mathcal{I}} \left| \left| V_I(\mathbf{x}^{(k)}) \right| - \frac{k}{n} \left| V_I(\mathbf{x}) \right| \right| \leq 4 + \sqrt{\frac{7n \log(n)}{2}} \ ,$$

with probability higher than $1 - 1/n^3$. Lemma B.19 is proved □

Proof of Lemma B.18. We first prove the upper bound

$$k d_\infty(\mathbf{x}, \mathbf{\Pi}_k) \leq 2\pi [\sup_{I \in \mathcal{I}} |V_I(\mathbf{x})| + 2] \ .$$

Recall that for a vector $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in \mathcal{C}^k$, we say that $\mathbf{x}$ is ordered, if these points are consecutive when one walks on the sphere with the trigonometric direction. Without loss of generality and for ease of exposition, we assume that the identity permutation is a latent order, that is $x_1, \ldots, x_k$ is ordered.

We define $\mathbf{x}_S^{**} = (x_1^{**}, \ldots, x_k^{**})$ a vector of $\mathbf{\Pi}_k$ as follows. The first point $x_1^{**} \in \mathcal{C}_k$ is a closest point to x_1^* with respect to d and the other points x_{j+1}^{**} are elements of $\mathcal{C}_k$ with arguments

$$\underline{x}_{j+1}^{**} = \underline{x}_1^{**} + j\frac{2\pi}{k} \pmod{2\pi}, \qquad \text{for } j = 1, \ldots, k-1 \ .$$

Fix any $i \in \{2, \dots, k\}$ and consider the intervals $I_i = [\underline{x}_1, \underline{x}_i]$ and $I'_i = [\underline{x}_1, \underline{x}_i^{**}]$. We have

$$d(x_i, x_i^{**}) \leq \big||I_i| - |I'_i|\big| \ . \tag{53}$$

Observe that $N_{I_i}(\mathbf{x}) = i$ since $x_1, \dots, x_k$ are ordered and all distinct. Hence,

$$\left|\frac{2\pi i}{k} - |I_i|\right| = \left|\frac{2\pi N_{I_i}}{k} - |I_i|\right| = 2\pi \frac{|V_{I_i}(\mathbf{x})|}{k} \leq 2\pi \sup_{I \in \mathcal{I}} \frac{|V_I(\mathbf{x})|}{k} \ . \tag{54}$$

Besides, we know that the length of I'_i is equal to $\big|[\underline{x}_1^{**}, \underline{x}_i^{**}]\big|$ up to an additional term $d(x_1, x_1^{**})$, that is

$$\Big||I'_i| - \big|[\underline{x}_1^{**}, \underline{x}_i^{**}]\big|\Big| \leq d(x_1, x_1^{**}) \ .$$

By construction of the x_j^{**}'s, we have $d(x_1, x_1^{**}) \leq 2\pi/k$ and $\big|[\underline{x}_1^{**}, \underline{x}_i^{**}]\big| = 2\pi(i-1)/k$. Hence, we obtain $k||I'_i| - 2\pi i| \leq 4\pi$. We then deduce from (54) and the triangular inequality that $k\big||I_i| - |I'_i|\big| \leq 2\pi \sup_{I \in \mathcal{I}} |V_I| + 4\pi$. Coming back to (53), taking the supremum over all $i \in \{2, \dots, k\}$, and noting that $d(x_1, x_1^{**}) \leq 2\pi/k$ leads us to

$$d_\infty(\mathbf{x}, \mathbf{x}^{**}) \leq \frac{2\pi}{k}\left[\sup_{I \in \mathcal{I}} |V_i| + 2\right] \ ,$$

where $\mathbf{x}^{**} \in \mathbf{\Pi}_k$. Finally, we take the minimum over $\mathbf{\Pi}_k$ to get the desired bound.

We now turn to the lower bound $kd_\infty(\mathbf{x}, \mathbf{\Pi}_k) \geq \pi[\sup_{I \in \mathcal{I}} |V_I(\mathbf{x})| - 4]$. Consider any such interval I and $\mathbf{x}^{**} \in \mathbf{\Pi}_k$. Since the entries of $\mathbf{x}^{**}$ are regularly spaced on $\mathcal{C}$, it follows that $|N_I(\mathbf{x}^{**}) - k|I|/(2\pi)| \leq 1$ so that $|V_I(\mathbf{x})| \leq |N_I(\mathbf{x}) - N_I(\mathbf{x}^{**})| + 1$. Now, assume that $N_I(\mathbf{x}) > N_I(\mathbf{x}^{**})$. We claim that $\sup_{j:\, x_j \in I} |x_j - x_j^{**}| \geq \frac{\pi}{k}[|N_I(\mathbf{x}) - N_I(\mathbf{x}^{**})| - 3]$. Otherwise, the set of x_j^{**} with j satisfying $x_j \in I$ is included in an interval of size

$$|I| + \frac{2\pi}{k}[N_I(\mathbf{x}) - N_I(\mathbf{x}^{**}) - 3] \leq \frac{2\pi}{k}[N_I(\mathbf{x}) - 2] \ .$$

This contradicts the fact that this set of equi-spaced points has size $N_I(\mathbf{x})$. If $N_I(\mathbf{x}) < N_I(\mathbf{x}^{**})$, we simply consider the complement[3] interval $\overline{I}$ that satisfies $N_{\overline{I}}(\mathbf{x}) = k - N_I(\mathbf{x})$ and $N_{\overline{I}}(\mathbf{x}^{**}) = k - N_I(\mathbf{x}^{**})$ to conclude that

$$\sup_{j:\, x_j \in \overline{I}} |x_j - x_j^{**}| \geq \frac{\pi}{k}\left[|N_{\overline{I}}(\mathbf{x}) - N_{\overline{I}}(\mathbf{x}^{**})| - 3\right] = \frac{\pi}{k}\left[|N_I(\mathbf{x}) - N_I(\mathbf{x}^{**})| - 3\right] \ .$$

Putting everything together, we have shown that

$$d_\infty(\mathbf{x}, \mathbf{x}^{**}) \geq \frac{\pi}{k}\left[|N_I(\mathbf{x}) - N_I(\mathbf{x}^{**})| - 3\right] \geq \frac{\pi}{k}\left[|V_I(\mathbf{x})| - 4\right] \ .$$

Taking the infimum over $\mathbf{x}^{**}$ and the supremum over I leads to the desired result. □

---

[3] Although $\overline{I}$ is an open interval, the arguments are still valid.

B.5. Proof of Corollary 3.3

Theorem 3.2 ensures that, conditionally to $\mathbf{x}^*$,

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}^*) \leq C'_{lLe}\left(d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right) \quad ,$$

with probability at least $1-5/n^2$. Thus, it suffices to show that, with probability at least $1-2/n^2$, one has

$$d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq C\sqrt{\frac{\log(n)}{n}} \quad ,$$

for some $C > 0$. We shall rely on Dvoretzky–Kiefer–Wolfowitz (DKW) inequality. Indeed, the arguments $\underline{x}_1^*, \ldots, \underline{x}_n^*$ are independent and uniformly distributed on $[0, 2\pi)$. Besides, any interval I of the torus $\mathbb{R}/(2\pi)$ can be represented as a union of at most two intervals of $[0, 2\pi)$. For any interval I, we denote $|I|$ its length and $N_I(\mathbf{x}^*)$ the number of points x_i^* whose argument lies in I. Then, we deduce from DKW inequality that, for any $t > 0$,

$$\mathbb{P}\left(\sup_{I\subset\mathbb{R}/(2\pi)}\left|\frac{N_I(\mathbf{x}^*)}{n} - \frac{|I|}{2\pi}\right| > 4t\right) \leq 2e^{-2nt^2} \quad .$$

We then choose $t = \sqrt{\log(n)/n}$ to obtain

$$\mathbb{P}\left(\sup_{I\subset\mathbb{R}/(2\pi)}\left|\frac{N_I(\mathbf{x}^*)}{n} - \frac{|I|}{2\pi}\right| > 4\sqrt{\frac{\log(n)}{n}}\right) \leq \frac{2}{n^2} \quad ,$$

Besides, by Lemma B.18, we know that the quantity $V_I(\mathbf{x}^*) = N_I(\mathbf{x}^*) - \frac{n|I|}{2\pi}$ introduced in (46) satisfies

$$d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq C\left(\sup_{I\subset\mathcal{I}}\frac{|V_I(\mathbf{x}^*)|}{n} + \frac{1}{n}\right) \quad ,$$

where $\mathcal{I}$ stands for the set of interval on the torus $\mathbb{R}/(2\pi)$. The last two displays lead to the desired result. □

Appendix C: Proof of the identifiability results and minimax lower bound

C.1. Proof of Proposition 2.2

For simplicity, we assume that $n/8$ is an integer in the rest of the example and we write $n = 8n_1$. The construction of f' mainly amounts to contracting the function f in some regions and dilating it in other regions which allows to contracting and dilating the positions $\mathbf{x}$.

Consider a partition of the latent space $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ in three arcs $\mathcal{C}_1 = (\underline{x}_n, \underline{x}_{n_1}] = (0, \pi/4]$, $\mathcal{C}_2 = (\underline{x}_{n_1}, \underline{x}_{4n_1}] = (\pi/4, \pi]$ and $\mathcal{C}_3 = (\underline{x}_{4n_1}, \underline{x}_n] = (\pi, 2\pi]$. For x and y belonging $\mathcal{C}_1$, define $f'_1(x,y)$ by $f'_1(x,y) = 1 - d(x,y)/\pi$. For $k = 1, \ldots, 2n_1$, define $x'_k = e^{\iota k\pi/n}$ and let $x'_n = x_n = 1$. In other words, we contract the positions x_k for $k = 1, \ldots, 2n_1$. Although we have not yet completely defined $\mathbf{x}'$, we already can certify that $\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}') \geq \pi/8$. Besides, we have $f'_1(x'_i, x'_j) = f(x_i, x_j)$ for all $i, j \in [2n_1] \cup \{n\}$.

For x and y in $\mathcal{C}_2$, we define $f'_2(x,y)$ by $f'_2(x,y) = 1 - d(x,y)/(3\pi)$. For $k = 1, \ldots, 2n_1$, we set $x'_{k+2n_1} = e^{\iota\pi/4}e^{\iota k3\pi/n}$. Again, observe that $f'_2(x'_i, x'_j) = f(x_i, x_j)$ for all integers i, j in $[2n_1 + 1, 4n_1]$. Finally, for x and y in $\mathcal{C}_3$, set $f'_3(x,y) = f(x,y)$, and let $x'_k = x_k$ for all integers $k = 4n_1 + 1, \ldots, n-1$. Obviously, we have $f'_3(x'_i, x'_j) = f(x_i, x_j)$ for all integers $i, j \in [4n_1, n]$.

It remains to deal with the situations where the pairs of points lie in different parts of the partition $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$. In the case where $x \in \mathcal{C}_1$ and $y \in \mathcal{C}_2$, define $f'_{1-2}(x,y) = f'_1(x, e^{\iota\pi/4}) + f'_2(e^{\iota\pi/4}, y) - 1$. For all integers $i \in [0, 2n_1]$ and $j \in [2n_1, 4n_1]$, we have already seen that $f'_1(x'_i, e^{\iota\pi/4}) = f(x_i, e^{\iota\pi/2})$ and $f'_1(e^{\iota\pi/4}, x'_j) = f(e^{\iota\pi/2}, x_j)$. Hence $f'_{1-2}(x'_i, x'_j) = f(x_i, e^{\iota\pi/2}) + f(e^{\iota\pi/2}, x_j) - 1 = f(x_i, x_j)$.

In the case where $x \in \mathcal{C}_1$ and $y \in \mathcal{C}_3$, define $f'_{1-3}(x,y) = f'_1(x, e^{\iota 0}) + f'_3(e^{\iota 0}, y) - 1$ if the length of the arc $[x, e^{\iota 0}] \cup (e^{\iota 0}, y]$ is less than π; otherwise, set $f'_{1-3}(x,y) = f'_1(x, e^{\iota\pi/4}) + f'_2(e^{\iota\pi/4}, e^{\iota\pi}) + f'_3(e^{\iota\pi}, y) - 2$. Since f admits similar decompositions, one can deduce from the above that $f'_{1-3}(x'_i, x'_j) = f(x_i, x_j)$ for all $i \in [1, 2n_1]$ and $j \in [4n_1 + 1, n]$.

The remaining cases can be handled in the same manner. Finally, we define the symmetric function f' on $\mathcal{C} \times \mathcal{C}$ relying on f'_1, f'_2, f'_3, f'_{1-2}, f'_{1-3}, and f'_{2-3}. Then, we can readily check that $f \in \mathcal{BL}[(3\pi)^{-1}, \pi^{-1}, 0]$ and that $f(x_i, x_j) = f'(x'_i, x'_j)$ for all $i, j \in [n]$. As a consequence, $(\mathbf{x}', f')$ belongs to $\mathcal{R}[F, (3\pi)^{-1}, \pi^{-1}, 0]$. One easily check that $\mathbf{x}' \in \mathcal{S}_{ev}$ and the result follows. ☐

C.2. Proof of (9) in Proposition 2.3

We show in the paragraph below that (9) is a consequence of the proof of Theorem 3.2 in the noiseless case ($E = 0$), after application of the triangular inequality. Indeed, since the noise is equal to zero, the conclusion of Theorem 3.2 is deterministic (and not with high probability anymore), so it can be used to prove deterministic inequalities such as (9). By doing so, we establish (9) via our localization algorithm (Theorem 3.2), though (9) is an approximation result (independent of any algorithm) which could be proved directly.

Consider any two representations $(\mathbf{x}, f)$ and $(\mathbf{x}', f')$ in $R[F, c_l, c_L, c_e]$ and apply our Localize-and-Refine procedure to noiseless observations $A = F$. The conclusion of Theorem 3.2 applies to both $\mathbf{x}$ and $\mathbf{x}'$, so that we have

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}) \quad \leq \quad C'_{lLe}\left(d_\infty(\mathbf{x}, \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right)$$

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}, Q\mathbf{x}') \quad\leq\quad C'_{lLe}\left(d_\infty(\mathbf{x}', \mathbf{\Pi}_n) + \sqrt{\frac{\log(n)}{n}}\right) \ .$$

Hence, it follows from the triangular inequality that

$$\begin{aligned}\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}') &= \min_{Q_1,\ Q_2\in\mathcal{O}} d_\infty(Q_1\mathbf{x}, Q_2\mathbf{x}') \\ &\leq C'_{lLe}\left(d_\infty(\mathbf{x}, \mathbf{\Pi}_n) + d_\infty(\mathbf{x}', \mathbf{\Pi}_n) + 2\sqrt{\frac{\log(n)}{n}}\right) \ .\end{aligned}$$

C.3. Proof of Theorem 5.1

We establish the lower bound $\sqrt{\log(n)/n}$ in the particular setting where the observations A_{ij} are independent Bernoulli random variables of parameters $F_{ij} = f_0(x_i, x_j)$, for the specific function

$$f_0(x_i, x_j) = (3/4) - d(x_i, x_j)/(4\pi), \tag{55}$$

with $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{\Pi}_n$. The corresponding probability distribution is denoted by $\mathbb{P}_{(\mathbf{x}, f_0)}$.

This minimax lower bound is based on Fano's method as stated below. For two configuration $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbf{\Pi}_n$, we denote the Kullback-Leibler divergence of $\mathbb{P}_{(\mathbf{x}, f_0)}$ and $\mathbb{P}_{(\mathbf{x}', f_0)}$ by $KL(\mathbb{P}_{(\mathbf{x}, f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}', f_0)})$. Besides, we quantify the quasi-metric $\rho(\mathbf{x}, \mathbf{x}') = \min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{x}')$. Given a radius $\delta > 0$ and a subset $\mathcal{S} \subset \mathbf{\Pi}_n$, the packing number $\mathcal{M}(\delta, \mathcal{S}', \rho)$ is defined as the largest number of points in $\mathcal{S}'$ that are at quasi-distance ρ at least δ away from each other. Below, we state a specific version of Fano's lemma.

Lemma C.1 (from [41])**.** *Consider any subset $\mathcal{S}' \subset \mathbf{\Pi}_n$. Define the Kullback-Leibler diameter of $\mathcal{S}'$ by*

$$d_{KL}(\mathcal{S}') = \sup_{\mathbf{x}, \mathbf{x}'\in\mathcal{S}'} KL(\mathbb{P}_{(\mathbf{x}, f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}', f_0)}) \ .$$

Then, for any estimator $\hat{\mathbf{x}}$ and for any $\delta > 0$, we have

$$\sup_{\mathbf{x}\in\mathcal{S}'} \quad \mathbb{P}_{(\mathbf{x}, f_0)}\left[\rho(\hat{\mathbf{x}}, \mathbf{x}) \geq \frac{\delta}{2}\right] \geq 1 - \frac{d_{KL}(\mathcal{S}') + \log(2)}{\log \mathcal{M}(\delta, \mathcal{S}', \rho)} \ .$$

In view of the above proposition, we mainly have to choose a suitable subset $\mathcal{S}'$, control its Kullback diameter, and get a sharp lower bound of its packing number. The main difficulty stems from the fact that the loss function $\rho(\mathbf{x}, \mathbf{y}) = \min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}, Q\mathbf{y})$ is a minimum over a collection of orthogonal transformations. It is therefore challenging to derive a tight lower bound for this loss.

Let $k := C'\sqrt{n\log(n)}$, for a small enough constant $C' \in (0, 1]$ that will be set later. Define $n/2$ vectors $\mathbf{x}^{(s)} \in \mathbf{\Pi}_n$, $s = 1, \ldots, n/2$, as follows. For each

$s \in [n/2]$, we define $x_j^{(s)}$ by its argument $\underline{x}_j^{(s)}$

$$\underline{x}_j^{(s)} = \frac{2\pi j}{n} \ , \qquad \forall j \in [n] \setminus \{s, s+k\} \ , \underline{x}_s^{(s)} = \frac{2\pi(s+k)}{n}, \quad \underline{x}_{s+k}^{(s)} = \frac{2\pi s}{n} \ .$$

Each vector of arguments $\underline{\mathbf{x}}^{(s)}$ is therefore equal to the vector $(2\pi j/n)_{j\in[n]}$ up to an exchange of the positions $2\pi s/n$ and $2\pi(s+k)/n$. This collection of $n/2$ vectors is denoted by $\mathcal{S}' := \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n/2)}\}$. Obviously $\mathcal{S}' \subset \mathbf{\Pi}_n$, and one can readily checks that

$$\rho(\mathbf{x}^{(t)}, \mathbf{x}^{(s)}) \geq \frac{\pi k}{n} \ , \qquad \forall s, t \in \left[\frac{n}{2}\right], \ s \neq t \ , \tag{56}$$

which in turn ensures that the packing number $\mathcal{M}(\delta_n, \mathcal{S}', \rho)$ of radius $\delta_n := \pi k/n$ satisfies $\mathcal{M}(\delta_n, \mathcal{S}', \rho) \geq n/2$. To upper bound the KL diameter of $\mathcal{S}'$, we use the following claim whose proof is postponed to the end of the section.

Claim C.2. *For any* $\mathbf{x}, \mathbf{x}' \in \mathcal{C}^n$, *we have*

$$KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) \leq 8 \sum_{i,j} (f_0(x_i, x_j) - f_0(x'_i, x'_j))^2 \ .$$

Together with the definition (55) of f_0, we get

$$KL(\mathbb{P}_{(\mathbf{x}^{(t)},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}^{(s)},f_0)}) \leq Cn\delta_n^2 \leq C(C')^2 \log(n) \ ,$$

for some numerical constant C. Then, choosing the constant C' in the definition of k such that $C' = (2\sqrt{C})^{-1}$ leads to $d_{KL}(\mathcal{S}') \leq \log(n)/4$.

Applying Lemma C.1 to this set $\mathcal{S}'$, we arrive at

$$\inf_{\hat{\mathbf{x}}} \sup_{\mathbf{x}\in\mathcal{S}'} \ \mathbb{P}_{(\mathbf{x},f_0)} \left[\rho(\hat{\mathbf{x}}, \mathbf{x}) \geq \frac{\delta_n}{2}\right] \geq 1 - \frac{\log(n)/4 + \log(2)}{\log(n/2)} \geq \frac{1}{2} \ ,$$

as soon as n is large enough. Theorem 5.1 is proved. □

Proof of Claim C.2. By definition of the Kullback-Leibler divergence, and $F_{ij} = f_0(x_i, x_j)$ and $F'_{ij} = f_0(x'_i, x'_j)$, we have

$$KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) = \sum_{i<j} F_{ij} \log \frac{F_{ij}}{F'_{ij}} + (1 - F_{ij}) \log \frac{1 - F_{ij}}{1 - F'_{ij}} \ ,$$

and since $\log(t) \leq t - 1$ for all $t > 0$, it follows that

$$KL(\mathbb{P}_{(\mathbf{x},f_0)} \,\|\, \mathbb{P}_{(\mathbf{x}',f_0)}) \leq \sum_{ij} \frac{(F_{ij} - F'_{ij})^2}{F'_{ij}(1 - F'_{ij})} \leq 8 \sum_{i,j} (F_{ij} - F'_{ij})^2 \ ,$$

where the second inequality follows from the fact that $1/4 \leq F'_{ij} \leq 3/4$. □

Appendix D: Proof for the spectral method

D.1. Proof of Theorem 4.1

Recall that the Spectral Localization (LS) algorithm is applied to the data matrix A_{SS}, where S is a subset of indices of $[n]$, with a cardinal number $|S| = n_0 = n/4$. We can assume that $S = \{1, \ldots, n_0\}$ for the ease of exposition. Vanilla Spectral Localization in LS algorithm returns $\hat{\mathbf{x}}_S^{\mathrm{VSA}} := (\hat{x}_1^{\mathrm{VSA}}, \ldots, \hat{x}_{n_0}^{\mathrm{VSA}})$ with $\hat{x}_i^{\mathrm{VSA}} := (\hat{x}_{ij}^{\mathrm{VSA}})_{j\in[2]} = \sqrt{\frac{n_0}{2}}(\hat{u}_i, \hat{v}_i) \in \mathbb{R}^2$, $i = 1, \ldots, n_0$. We denote by $\hat{\lambda}_0^{(S)} \geq \ldots \geq \hat{\lambda}_{n_0-1}^{(S)}$ the eigenvalues of the adjacency matrix A_{SS}.

Note that the position estimates $\hat{x}_i^{\mathrm{VSA}}$ do not lie on the unit sphere $\mathcal{C}$. As a consequence, the quantity $d(\hat{x}_i^{\mathrm{VSA}}, x_i^*)$ is not defined, and we will use the distance $\|\hat{\mathbf{x}}_S^{\mathrm{VSA}} - \mathbf{x}_S^*\|_1 = \sum_{i=1}^{n_0}\sum_{j=1}^{2}|\hat{x}_{ij}^{\mathrm{VSA}} - x_{ij}^*|$ where $\mathbf{x}_S^*$ is interpreted as a $2 \times n_0$ matrix. Besides, since $\mathbf{x}_S^*$ can only be recovered up to orthogonal transformations, we consider the loss $\min\limits_{Q\in\mathcal{O}}\|\hat{\mathbf{x}}_S^{\mathrm{VSA}} - Q\mathbf{x}_S^*\|_1$ where Q is interpreted as 2×2 orthogonal matrix.

Let $\lambda_0^{*(S)} \geq \ldots \geq \lambda_{n_0-1}^{*(S)}$ denote the eigenvalues of the signal matrix $F_{SS} := [f(x_i^*, x_j^*))]_{i,j\in S}$. We denote by $\Delta_1^{(S)} := \lambda_0^{*(S)} - \lambda_1^{*(S)}$ and $\Delta_2^{(S)} := \lambda_2^{*(S)} - \lambda_3^{*(S)}$ the two relevant spectral gaps.

Proposition D.1. *Let $n_0 \geq 4$, and f be a geometric function as defined in (23), such that f belongs to $\mathcal{BL}[c_l, c_L, c_e]$. Let $c_a > 0$ be any positive constant. Assume that the latent positions $\mathbf{x}_S^*$ fulfill the following inequality*

$$d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) \leq c_a\sqrt{\frac{\log(n)}{n}}\ . \tag{57}$$

Then, with probability higher than $1 - 1/n^2$, the spectral estimator $\hat{\mathbf{x}}_S^{VSA}$ satisfies

$$\min_{Q\in\mathcal{O}}\|\hat{\mathbf{x}}_S^{VSA} - Q\mathbf{x}_S^*\|_1 \leq C_{lLea}\,\frac{n\sqrt{n\log(n)}}{\left(\Delta_1^{(S)} \wedge \Delta_2^{(S)}\right) \vee 1}\ .$$

Proposition D.1 is based on the fact that the signal matrix F_{SS} is well approximated by a circulant and circular-R matrix, which benefits from nice spectral properties. See Appendix D.4 for a proof.

Assumption (24) of the theorem states that $d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq c_a\sqrt{\log(n)/n}$. Since Lemma B.17 ensures that $d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) \leq C[d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) + \sqrt{\log(n)/n}]$ with probability higher than $1 - 1/n^2$, we get the bound

$$d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) \leq C_a\ \sqrt{\log(n)/n}\ , \tag{58}$$

which holds with probability higher than $1 - 1/n^2$.

The ℓ^1-type localization bound in (D.1) depends on the spectral gap $\Delta_1^{(S)} \wedge \Delta_2^{(S)}$ of the signal matrix F_{SS}. We combine the next lemma with the assumption $\Delta_1 \wedge \Delta_2 \geq c_b n$ of the theorem to get the following lower bound

$$\Delta_1^{(S)} \wedge \Delta_2^{(S)} \geq c_b n/4\ , \tag{59}$$

which holds with probability higher than $1-1/n^2$, as soon as $n \geq C_{abL}$.

Lemma D.2. *If $\Delta_1 \wedge \Delta_2 \geq c_b n$ for some constant $c_b > 0$, then with probability higher than $1-1/n^2$ we have $\Delta_1^{(S)} \wedge \Delta_2^{(S)} \geq c_b n/4$ for all $n \geq C_{ab}$ where C_{abL} is a positive quantity depending only on c_a, c_b, and c_L.*

Hence, By (59) and (58) the conditions of Proposition D.1 are satisfied. In summary, there exists an event of probability higher than $1-3/n^2$ such that

$$\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{\mathrm{VSA}} - Q\mathbf{x}_S^*\|_1 \leq C_{lLeab}\sqrt{n\log(n)}\ , \tag{60}$$

as soon as $n \geq C_{abL}$. For $n \leq C_{abL}$, the bound (60) trivially holds provided that we adjust the constant C_{lLeab} if necessary.

Since $\hat{\mathbf{x}}_S^{\mathrm{VSA}}$ does not lie in $\mathbf{\Pi}_{n_0} \subset \mathcal{C}^{n_0}$, we cannot directly plug it into the local refinement step defined by (16). Accordingly, the Uniform Approximation (UA) in LS algorithm, projects $\hat{\mathbf{x}}_S^{\mathrm{VSA}}$ onto $\mathbf{\Pi}_{n_0} \subset \mathcal{C}^{n_0}$. The UA outputs a vector $\tilde{\mathbf{x}}_S^{(1)}$ in $\mathbf{\Pi}_{n_0}$ that is close to the input $\hat{\mathbf{x}}_S^{\mathrm{VSA}}$ – see Lemma D.3 below.

Lemma D.3 actually gives a more general result that holds for any input in $\mathbb{R}^{2\times n_0}$ given to UA. For clarity, we write below the UA procedure in full generality.

```
Uniform Approximation (UA) in $\mathbf{\Pi}_{n_0}$
Input: $\mathbf{x}_S = (x_1, \dots, x_{n_0}) \in \mathbb{R}^{2\times n_0}$.
  1. Set $z_i = x_i/\|x_i\|_2$, for $i = 1, \dots, n_0$.
  2. Pick any permutation $\sigma$ such that $z_{\sigma(1)}, \dots, z_{\sigma(n_0)}$ is in trigonometric order.
  3. Set $\tilde{x}_{\sigma(i)} = e^{\iota\frac{2\pi(\hat{k}+i)}{n_0}}$,   where   $\widehat{k} \in \mathrm{argmin}_{k\in[n_0]} \sum_{i=1}^{n_0} \left\| e^{\iota\frac{2\pi(k+i)}{n_0}} - z_{\sigma(i)} \right\|_1$.
Output: $\tilde{\mathbf{x}}_S \in \mathbf{\Pi}_{n_0}$.
```

Lemma D.3. *Let $\mathbf{x}_S^* \in \mathcal{C}^{n_0}$. For any input $\mathbf{x}_S \in \mathbb{R}^{2\times n_0}$, UA returns a vector $\widetilde{\mathbf{x}}_S \in \mathbf{\Pi}_{n_0}$ such that*

$$\min_{Q\in\mathcal{O}} \|\widetilde{\mathbf{x}}_S - Q\mathbf{x}_S^*\|_1 \lesssim \min_{Q\in\mathcal{O}} \|\mathbf{x}_S - Q\mathbf{x}_S^*\|_1 + n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + 1\ .$$

By (58–60) and Lemma D.3, the projection $\widetilde{\mathbf{x}}_S^{(1)}$ satisfies

$$\min_{Q\in\mathcal{O}} \|\widetilde{\mathbf{x}}_S^{(1)} - Q\mathbf{x}_S^*\|_1 \lesssim \min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{\mathrm{VSA}} - Q\mathbf{x}_S^*\|_1 + n d_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + 1 \leq C_{lLeab}\sqrt{n\log(n)} \tag{61}$$

with probability higher than $1-3/n^2$.

Finally, we plug $\widetilde{\mathbf{x}}_S^{(1)}$ in the criterion (16) to localize the remaining points. In other words, we compute, for $i \in \overline{S}$,

$$\hat{x}_i^{(2)} = \operatorname*{argmin}_{z\in\mathcal{C}_{n_0}} \langle A_{i,S}, D(z, \tilde{\mathbf{x}}_S^{(1)})\rangle\ , \tag{62}$$

and get the position estimates $\hat{\mathbf{x}}^{(2)}_{\bar{S}} = [\hat{x}^{(2)}_i]_{i\in \bar{S}}$. As a direct consequence of (61) and Propositions 3.5 (and the equivalence between the ℓ^1 norm in $\mathbb{R}^2$ and the distance d_1 in the sphere $\mathcal{C}$), we arrive at the following uniform bound

$$\min_{Q\in\mathcal{O}} d_\infty(\hat{\mathbf{x}}^{(2)}_{\bar{S}}, Q\mathbf{x}^*_{\bar{S}}) \leq C_{lLeab}\sqrt{\frac{\log(n)}{n}}\ ,$$

which holds with probability higher than $1-4/n^2$.

As in section 3.4, we finally rely on a cross-validation scheme to estimate and realign all the positions. This straightforwardly allows us to uniformly localize, with probability higher than $1-9/n^2$, all positions $\mathbf{x}^*_i$ within an error of the order of $\sqrt{\log(n)/n}$. This concludes the proof of Theorem 4.1. ☐

D.2. Proof of Lemma D.2

Recall that $S=[n_0]$ for the ease of exposition. In order too show that the spectrums of F and $F_{SS} := [f(x^*_i, x^*_j)]_{i,j\in S}$ are linked together, we introduce an intermediate matrix $F^{(4)} := [f(x^{*(4)}_i, x^{*(4)}_j)]_{i,j\in[n]}$ based on the vector $\mathbf{x}^{*(4)}_S \in \mathcal{C}^n$ with coordinates $x^{*(4)}_i = x^*_{\lceil i/4\rceil}$ for $i\in[n_0]$ (where $\lceil\cdot\rceil$ denotes the ceiling function). In other words, we replicate 4-times each coordinate of the vector $\mathbf{x}^*_S$ to get the vector $\mathbf{x}^{*(4)}_S$ of size n which is close to $\mathbf{x}^*$.

Let us show first that the spectrums of F_{SS} and $F^{(4)}$ are almost the same. By construction of $F^{(4)}$, each of the n_0 eigenvectors of F_{SS} can be transformed into an eigenvector of $F^{(4)}$, by replicating 4-times the coordinates of these vectors. Besides, the rank of $F^{(4)}$ is the same as that of F_{SS}. We deduce that all non-zero eigenvalues of $F^{(4)}$ are eigenvalues of $2F_{SS}$. Formally, denoting the eigenvalues of $F^{(4)}$ by $\lambda'_0 \geq \ldots \geq \lambda'_{n-1}$, and recalling that the eigenvalues of F_{SS} are denoted by $\lambda^{*(S)}_0 \geq \ldots \geq \lambda^{*(S)}_{n_0-1}$, we have

$$2\lambda^{*(S)}_i = \lambda'_i\ , \qquad \text{for } i=1,\ldots,n_0\ . \tag{63}$$

We then show that the spectrums of $F^{(4)}$ and F are close. By (58) there is a probability higher than $1-1/n^2$ that $d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) \leq C_a\sqrt{\log(n)/n}$. Hence, one can readily check that

$$d_\infty(\mathbf{x}^{*(4)}, \mathbf{\Pi}_n) \leq d_\infty(\mathbf{x}^{*(4)}, \mathbf{\Pi}_{n_0}) + 2\pi/n_0 \leq C'_a\sqrt{\log(n)/n}\ .$$

Furthermore, Assumption (24) of the theorem ensures that

$$d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq c_a\sqrt{\log(n)/n}\ .$$

Therefore, both $\mathbf{x}^{*(4)}$ and $\mathbf{x}^*$ are close to $\mathbf{\Pi}_n$. Since (any) two elements of $\mathbf{\Pi}_n$ are equal up to a permutation of their indices $[n]$, we deduce that there exists a permutation σ of $[n]$ satisfying $d_\infty(\mathbf{x}^{*(4)}, \mathbf{x}^*_\sigma) \leq C_a\sqrt{\log(n)/n}$. Combining this with the bi-Lipschitz condition (5) we deduce that

$$\|F^{(4)} - F_\sigma\|_2 \leq C_a c_L\sqrt{n\log(n)}\ .$$

We are now ready to control the difference between the spectrums of F and $F^{(4)}$. Recalling that $\lambda_0^* \geq \ldots \geq \lambda_{n-1}^*$ denote the eigenvalues of F, and since F_σ has the same eigenvalues as F, it follows from Weyl's inequality (see e.g. [39, page 45]) that

$$|\lambda_i' - \lambda_i^*| \leq \|F^{(4)} - F_\sigma\|_{op} \leq \|F^{(4)} - F_\sigma\|_2 \leq C_{aL}\sqrt{n\log(n)}\ , \tag{64}$$

for all $i = 0, \ldots, n-1$, and some constant C_{aL} depending only on c_a and c_L.

Gathering (63–64), we conclude that the following implication holds. If $\Delta_1 := \lambda_0^* - \lambda_1^*$ and $\Delta_2 := \lambda_2^* - \lambda_3^*$ satisfies $\Delta \wedge \Delta_2 \geq c_b n$, then $\Delta_1^{(S)} := \lambda_0^{*(S)} - \lambda_1^{*(S)}$ and $\Delta_2^{(S)} := \lambda_2^{*(S)} - \lambda_3^{*(S)}$ fulfills $\Delta_1^{(S)} \wedge \Delta_2^{(S)} \geq \frac{c_b}{2}n - C\sqrt{n\log(n)}$ which is larger than $c_b n/4$ as soon as $n \geq n_{ab}$ for n_{ab} the smallest integer satisfying $c_b n \geq 4C_{aL}\sqrt{n\log(n)}$. □

D.3. Proof of Lemma D.3 (Uniform approximation)

Recall that for a vector $\mathbf{x} = (x_1, x_2, \ldots, x_{n_0}) \in \mathcal{C}^{n_0}$, we say that $\mathbf{x}$ is ordered, if these points are consecutive when one walks on the circle following the trigonometric direction.

We introduce some notation. For any vector $\mathbf{v} = (v_1, \ldots, v_{n_0}) \in \mathcal{C}^{n_0}$, denote $\mathbf{\Pi}_{n_0}(\mathbf{v})$ all elements $\mathbf{u} = (u_1, \ldots, u_{n_0})$ of $\mathbf{\Pi}_{n_0}$ such that, for all permutations σ making $v_{\sigma(1)}, \ldots, v_{\sigma(n_0)}$ ordered, the sequence $u_{\sigma(1)}, \ldots, u_{\sigma(n_0)}$ is ordered. For any vector $\mathbf{v} \in \mathcal{C}^{n_0}$ with distinct values, the set $\mathbf{\Pi}_{n_0}(\mathbf{v})$ can be described by a single element $\mathbf{u} \in \mathbf{\Pi}_{n_0}(\mathbf{v})$ and all circular permutations of $\mathbf{u}$.

The bound of Lemma D.3 trivially holds for $n_0 \leq 3$. Henceforth, we assume that $n_0 \geq 4$. The next lemma is a key element in the proof; it states that reordering two vectors is almost optimal for minimizing their d_1 distance. This result is fairly classical for real vectors. Here, as the vectors $\mathbf{v}$ and $\mathbf{u}$ take their values on $\mathcal{C}$ the proof is slightly more complicated.

Lemma D.4. *Consider any $\mathbf{v} \in \mathcal{C}^{n_0}$. Provided that $n_0 \geq 4$, we have*

$$\min_{\mathbf{u} \in \mathbf{\Pi}_{n_0}(\mathbf{v})} d_1(\mathbf{v}, \mathbf{u}_\sigma) \lesssim \min_{\mathbf{u} \in \mathbf{\Pi}_{n_0}} d_1(\mathbf{v}, \mathbf{u})\ .$$

Recall that $S = \{1, \ldots, n_0\}$ for the ease of exposition. We shall prove the following statement which implies Lemma D.3. For any $\mathbf{x}_S^* \in \mathcal{C}^{n_0}$, any input $\mathbf{x}_S \in \mathbb{R}^{2\times n_0}$, and any $Q \in \mathcal{O}$, the vector $\widetilde{\mathbf{x}}_S$ of $\mathbf{\Pi}_{n_0}$ fulfills

$$\|\widetilde{\mathbf{x}}_S - Q\mathbf{x}_S^*\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}_S^*\|_1 + nd_\infty(\mathbf{x}_S^*, \mathbf{\Pi}_{n_0}) + 1. \tag{65}$$

UA computes in step 1 the projection $\mathbf{z}_S = [x_i/\|x_i\|_2]_{i\in S}$ of the input $\mathbf{x}_S$ onto $\mathcal{C}^{n_0}$. Given the projection $\mathbf{z}_S$, UA picks in step 3 a vector $\tilde{\mathbf{x}}_S \in \mathbf{\Pi}_{n_0}(\mathbf{z}_S)$ that has the smallest ℓ^1-error:

$$\tilde{\mathbf{x}}_S \in \operatorname*{argmin}_{\mathbf{u} \in \mathbf{\Pi}_{n_0}(\mathbf{z}_S)} \|\mathbf{z}_S - \mathbf{u}\|_1.$$

It follows from these definitions and the equivalence between the distance d on $\mathcal{C}$ and ℓ^1-norm in $\mathbb{R}^2$ that

$$\|\mathbf{z}_S - \tilde{\mathbf{x}}_S\|_1 = \min_{\mathbf{v}\in\mathbf{\Pi}_{n_0}(\mathbf{z}_S)} \|\mathbf{z}_S - \mathbf{v}\|_1 \lesssim \min_{\mathbf{v}\in\mathbf{\Pi}_{n_0}(\mathbf{z}_S)} d_1(\mathbf{z}_S, \mathbf{v}) \ .$$

Gathering this bound with Lemma D.4, we derive that

$$\|\mathbf{z}_S - \tilde{\mathbf{x}}_S\|_1 \lesssim \min_{\mathbf{u}\in\mathbf{\Pi}_{n_0}} d_1(\mathbf{z}_S, \mathbf{u}) \lesssim \min_{\mathbf{u}\in\mathbf{\Pi}_{n_0}} \|\mathbf{z}_S - \mathbf{u}\|_1 \ .$$

As a consequence, it suffices to exhibit some $u \in \mathbf{\Pi}_{n_0}$ such that its ℓ^1 distance to the projection $\mathbf{z}_S$ is small. This is precisely the purpose of the next lemma.

Lemma D.5. *Consider any matrix $Q \in \mathcal{O}$. There exists $\mathbf{y} \in \mathbf{\Pi}_{n_0}$ such that*

$$\|\mathbf{z}_S - \mathbf{y}\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}^*_S\|_1 + n d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + 1 \ .$$

We conclude that

$$\|\mathbf{z}_S - \tilde{\mathbf{x}}_S\|_1 \lesssim \|\mathbf{z}_S - \mathbf{y}\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}^*_S\|_1 + n d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + 1 \ .$$

By triangular inequality, we have

$$\|\tilde{\mathbf{x}}_S - \mathbf{x}_S\|_1 \leq \|\tilde{\mathbf{x}}_S - \mathbf{z}_S\|_1 + \|\mathbf{z}_S - \mathbf{x}_S\|_1 \ .$$

The definition of a projection –and the equivalence between the ℓ^1-norm and the euclidean norm in $\mathbb{R}^2$– ensure that

$$\|\mathbf{z}_S - \mathbf{x}_S\|_1 \lesssim \|Q\mathbf{x}^*_S - \mathbf{x}_S\|_1 \ ,$$

since $\mathbf{z}_S$ is the projection of $\mathbf{x}_S$ on $\mathcal{C}^{n_0}$ and $Q\mathbf{x}^*_S$ is an element of $\mathcal{C}^{n_0}$. The last three displays allow us to conclude that

$$\|\tilde{\mathbf{x}}_S - \mathbf{x}_S\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}^*_S\|_1 + n d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0}) + 1 \ ,$$

which gives (65) using the triangle inequality again. □

D.3.1. Proofs of Lemma D.5

Let $\mathbf{x}^{**}_S \in \mathbf{\Pi}_{n_0}$ be a closest approximation of $\mathbf{x}^*_S$ in $\mathbf{\Pi}_{n_0}$, that is, such that $d_\infty(\mathbf{x}^*_S, \mathbf{x}^{**}_S) = d_\infty(\mathbf{x}^*_S, \mathbf{\Pi}_{n_0})$. The triangular inequality gives

$$\|Q\mathbf{x}^{**}_S - \mathbf{z}_S\|_1 \leq \|Q\mathbf{x}^{**}_S - \mathbf{x}_S\|_1 + \|\mathbf{x}_S - \mathbf{z}_S\|_1 \lesssim \|Q\mathbf{x}^{**}_S - \mathbf{x}_S\|_1 \ ,$$

where the last inequality comes from the definition of a projection and the equivalence between the ℓ^1-norm and the euclidean norm in $\mathbb{R}^2$. By the triangular inequality again, we get

$$\|Q\mathbf{x}^{**}_S - \mathbf{x}_S\|_1 \leq \|Q\mathbf{x}^{**}_S - Q\mathbf{x}^*_S\|_1 + \|Q\mathbf{x}^*_S - \mathbf{x}_S\|_1 \ .$$

An orthogonal transformation preserves the distances.

$$\|Q\mathbf{x}_S^{**} - Q\mathbf{x}_S^{*}\|_1 = \|\mathbf{x}_S^{**} - \mathbf{x}_S^{*}\|_1 \lesssim d_1(\mathbf{x}_S^{**}, \mathbf{x}_S^{*}) \leq n d_\infty(\mathbf{x}_S^{*}, \mathbf{\Pi}_{n_0}) \ ,$$

where we use again the equivalence between the distance d in $\mathcal{C}$ and the ℓ^1-norm in $\mathbb{R}^2$. Putting everything together, we conclude that

$$\|Q\mathbf{x}_S^{**} - \mathbf{z}_S\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}_S^{*}\|_1 + n d_\infty(\mathbf{x}_S^{*}, \mathbf{\Pi}_{n_0}) \ .$$

Although $\mathbf{x}_S^{**}$ belongs to $\mathbf{\Pi}_{n_0}$, this is not necessarily the case for $Q\mathbf{x}_S^{**}$. Nevertheless, it is easy to check that there exists some $Q' \in \mathcal{O}$ such that $Q'\mathbf{x}_S^{**} \in \mathbf{\Pi}_{n_0}$ and $\|Q'\mathbf{x}_S^{**} - Q\mathbf{x}_S^{**}\|_1 \lesssim 1$. Setting $\mathbf{y} := Q'\mathbf{x}_S^{**} \in \mathbf{\Pi}_{n_0}$, then we see that

$$\|\mathbf{y} - \mathbf{z}_S\|_1 \lesssim \|\mathbf{x}_S - Q\mathbf{x}_S^{*}\|_1 + n d_\infty(\mathbf{x}_S^{*}, \mathbf{\Pi}_{n_0}) + 1 \ , \tag{66}$$

which concludes the proof. □

D.3.2. Proof of Lemma D.4

Fix any vector $\mathbf{v} \in \mathcal{C}^{n_0}$ and any $\mathbf{u} \in \mathbf{\Pi}_{n_0}$. We shall prove that

$$\min_{\mathbf{w} \in \mathbf{\Pi}_{n_0}(\mathbf{v})} d_1(\mathbf{v}, \mathbf{w}) \lesssim d_1(\mathbf{v}, \mathbf{u}) \ .$$

Let τ be a permutation ordering the coordinates of $\mathbf{v}$ on the unit sphere, meaning that $v_{\tau(1)}, \ldots v_{\tau(n_0)}$ is ordered. For simplicity and without loss of generality, assume that τ is the identity. Since $\mathbf{u} \in \mathbf{\Pi}_{n_0}$, it there suffices to prove the existence of a permutation σ of $[n_0]$ such that u_σ is ordered and

$$d_1(\mathbf{v}, \mathbf{u}_\sigma) \lesssim d_1(\mathbf{v}, \mathbf{u})$$

Define the set of 'bad' indices $\mathcal{B} = \{i : d(u_i, v_i) \geq \pi/16\}$. If the cardinal of $\mathcal{B}$ is larger than $n_0/2$, then $d_1(\mathbf{u}, \mathbf{v}) \geq n_0\pi/32$ and any vector $\mathbf{u}_\sigma$ satisfies $d_1(\mathbf{v}, \mathbf{u}_\sigma) \leq n_0\pi \leq 32 d_1(\mathbf{v}, \mathbf{u})$. Hence, we assume henceforth that $|\mathcal{B}| \leq n_0/2$. First, we focus on the set of 'good' indices $\mathcal{G} = [n_0] \backslash \mathcal{B}$. We establish the following claim at the end of the proof.

Claim D.6. *There exists a permutation σ of $\mathcal{G}$ such that the sequence $(u_{\sigma(j)})$ with $j \in \mathcal{G}$ is ordered and*

$$\sum_{i \in \mathcal{G}} d(u_{\sigma(i)}, v_i) \leq \sum_{i \in \mathcal{G}} d(u_i, v_i)$$

Hence, it is possible to order the restriction of u to $\mathcal{G}$ without increasing the sum of the distances. It remains to transform σ into a permutation of $[n_0]$. We iteratively add elements of $\mathcal{B}$ into σ. Consider any $i \in \mathcal{B}$. Let k and l be the two consecutive (modulo n_0) elements of $\mathcal{G}$ such that $\underline{u}_i$ belongs to the interval $[\underline{u}_{\sigma(k)}, \underline{u}_{\sigma(l)})$ of the torus $\mathbb{R}/(2\pi)$. Let r and s be the two consecutive elements of $\mathcal{G}$ such that $i \in (r,s)$ (where we work modulo n_0). Then, we define the permutation σ' of $(\mathcal{G}\cup\{i\})$ as follows.

If $(r,s)=(k,l)$, then we take $\sigma'(j)=\sigma(j)$ if $j\in\mathcal{G}$ and $\sigma'(i)=i$. One readily checks that the sequence $(u_{\sigma'(j)})$ with $j\in\mathcal{G}\cup\{i\}$ is ordered and that $\sum_{j\in\mathcal{G}\cup\{i\}} d(u_{\sigma'(j)},v_j)\leq \sum_{j\in\mathcal{G}\cup\{i\}} d(u_j,v_j)$.

Otherwise, we set $\sigma'(i)=\sigma(s)$ and $\sigma'(k)=i$. For $j\in\mathcal{G}$, let $\mathrm{succ}_{\mathcal{G}}(j)$ be the successor of $j\in\mathcal{G}$, that is the smallest index $j'\in\mathcal{G}$ which is larger than j (modulo n_0). For any $j\in\mathcal{G}$ in the segment $[s,k)$, we set $\sigma'(j)=\sigma(\mathrm{succ}_{\mathcal{G}}(j))$. Besides, we set $\sigma'(j)=\sigma(j)$ for all $j\in\mathcal{G}$ in the segment $[l,r]$. In other words, we have shifted all elements in the segment $[s,k]$ to successfully include i in the permutation σ'. It follows from this definition that the sequence $u_{\sigma'(j)}$ with $j\in\mathcal{G}\cup\{i\}$ is ordered. By the triangular inequality, we have

$$\begin{aligned}
\sum_{j\in\mathcal{G}\cup\{i\}} d(u_{\sigma'(j)},v_j) &= \sum_{j\in\mathcal{G}\cap[l,r]} d(u_{\sigma(j)},v_j)+d(u_{\sigma(s)},v_i)+d(u_i,v_k)\\
&+\sum_{j\in\mathcal{G}\cap[s,k)} d(u_{\sigma(\mathrm{succ}_{\mathcal{G}}(j))},v_j)\\
&\leq 2\pi+\sum_{j\in\mathcal{G}} d(u_{\sigma(j)},v_j)+\sum_{j\in\mathcal{G}\cap[s,k)} d(u_{\sigma(j)},u_{\sigma(\mathrm{succ}_{\mathcal{G}}(j))})\\
&\leq 4\pi+\sum_{j\in\mathcal{G}} d(u_{\sigma(j)},v_j)\leq 4\pi+\sum_{j\in\mathcal{G}\cup\{i\}} d(u_j,v_j)\ ,
\end{aligned}$$

where we used in the third line that $\sum_{j\in\mathcal{G}\cap[s,k)} d(u_{\sigma(j)},u_{\sigma(\mathrm{succ}_{\mathcal{G}}(j))})\leq 2\pi$. Indeed, the sequence $(u_{\sigma(j)})_{j\in\mathcal{G}\cap[s,k)}$ is ordered on the sphere and this sum is therefore equal to the length of the arc $[\underline{u}_{\sigma(s)},\underline{u}_{\sigma(k)}]$.

By a straightforward induction, we manage to build a permutation $\overline{\sigma}$ on $[n_0]$ such that $(u_{\overline{\sigma}(j)})_{j\in[n_0]}$ is ordered and

$$\begin{aligned}
\sum_{j\in[n_0]} d(u_{\overline{\sigma}(j)},v_j) &\leq 4\pi\left|\left\{i\in[n_0]\,:\,d(u_i,v_i)\geq\frac{\pi}{16}\right\}\right|+\sum_{j\in[n_0]} d(u_j,v_j)\\
&\leq 65\sum_{j\in[n_0]} d(u_j,v_j)\ ,
\end{aligned}$$

where we used Markov's inequality in the last line. We have shown the desired result. □

Proof of claim D.6. Without loss of generality, we assume in the proof that $\mathcal{B}=\emptyset$ so that we build a permutation σ of $[n_0]$. Since $\mathcal{B}=\emptyset$, $\mathbf{u}\in\mathbf{\Pi}_{n_0}$ satisfies $d_\infty(\mathbf{u},\mathbf{v})\leq\pi/16$. We shall iteratively build a permutation σ such that $\mathbf{u}_\sigma$ is ordered. Let us first partition the one-dimensional torus $\mathbb{R}/(2\pi)$ into three parts $\mathbb{R}/(2\pi)=\mathcal{D}_1\cup\mathcal{D}_2\cup\mathcal{D}_3$ where $\mathcal{D}_s=\left[(s-1)\frac{2\pi}{3},s\frac{2\pi}{3}\right)$ for $s=1,2,3$.

For $s=1,2,3$, define $I_s=\{i:\underline{v}_i\in\mathcal{D}_s\}$. Since $d_\infty(\mathbf{u},\mathbf{v})\leq\pi/16$, it follows that $\{\underline{u}_i:i\in I_s\}\subset\left[(s-1)\frac{2\pi}{3}-\frac{\pi}{16},s\frac{2\pi}{3}+\frac{\pi}{16}\right)=\mathcal{D}'_s$. Note that the diameter of

$\mathcal{D}'_s$ is smaller $2\pi/3 + \pi/8 < \pi$. We have the decomposition

$$d_1(\mathbf{v}, \mathbf{u}) = \sum_{s=1}^{3} \sum_{i \in I_s} d(v_i, u_i) \ .$$

For $s = 1, 2, 3$, let σ_s denote the permutation of I_s such that the sequence $u_{\sigma_s(i)}$ is ordered when i is in I_s. Since the diameter of $\mathcal{D}'_s$ is at most π, the sequence $\underline{u}_{\sigma_s(i)}$ in $\mathcal{D}'_s$ is isometric to an increasing sequence of points in $[0, \pi] \subset \mathbb{R}$ endowed with the absolute value distance. It goes the same for the ordered sequence $\underline{v}_i$ in $\mathcal{D}'_s$. Next, we use the following classical property.

Claim D.7. *Let $l \geq 1$ be an integer and $\mathbf{a}, \mathbf{b}$ be two monotonic vectors of $\mathbb{R}^l$, that is, $a_1 \leq a_2 \leq \ldots \leq a_l$ and $b_1 \leq b_2 \leq \ldots \leq b_l$. Then, for any permutation τ of the indices $\{1, \ldots, l\}$, we have*

$$\sum_{j=1}^{l} |a_i - b_i| \leq \sum_{j=1}^{l} |a_i - b_{\tau(i)}| \quad \text{and} \quad \max(|a_i - b_i|) \leq \max(|a_i - b_{\tau(i)}|) \ .$$

It follows that, for $s = 1, 2, 3$, we have

$$\sum_{i \in I_s} d(v_i, u_{\sigma_s(i)}) \leq \sum_{i \in I_s} d(v_i, u_i) \ .$$

Let σ be the permutation such that $\sigma(i) = \sigma_s(i)$ if $i \in I_s$. Obviously, we have $d_1(\mathbf{v}, \mathbf{u}_\sigma) \leq d_1(\mathbf{v}, \mathbf{u})$. Besides, u_σ is ordered except possibly at the indices $J_s = \{i : \underline{u}_{\sigma(i)} \in [(s-1)\frac{2\pi}{3} - \frac{\pi}{16}; (s-1)\frac{2\pi}{3} + \frac{\pi}{16}]\}$ with $s = 1, 2, 3$. Since $\max_i d(u_{\sigma(i)}, v_i) \leq \frac{\pi}{16}$ by the second part of the above claim, all $\underline{u}_{\sigma(i)}$ and $\underline{v}_i$ with $i \in J_s$ belong to an interval of length smaller than π. Besides,

$$d_1(\mathbf{v}, \mathbf{u}_\sigma) = \sum_{j \notin (\cup_s J_s)} d(v_i, u_{\sigma_s(i)}) + \sum_{s=1}^{3} \sum_{i \in J_s} d(v_i, u_{\sigma(i)}) \ .$$

Hence, we can build as previously partitions σ'_s of J_s that make $u_{\sigma'_s(\sigma(i))}$ ordered on J_s and so that

$$\sum_{i \in J_s} d(v_i, u_{\sigma'_s(\sigma(i))}) \leq \sum_{i \in J_s} d(v_i, u_{\sigma(i)}) \ .$$

Defining $\overline{\sigma}(i) = \sigma'_s(\sigma(i))$ if $i \in J_s$ for $s = 1, 2, 3$ and $\overline{\sigma}(i) = \sigma(i)$ otherwise, we conclude that $\mathbf{u}_{\overline{\sigma}}$ is ordered and that $d_1(\mathbf{v}, \mathbf{u}_{\overline{\sigma}}) \leq d_1(\mathbf{v}, \mathbf{u})$. □

D.4. Proof of Proposition D.1

Under the extra assumption that f is geometric, i.e. f satisfies (23), we will show that the estimation error of the spectral algorithm is bounded by $\frac{n\sqrt{n\log(n)}}{\left(\Delta_1^{(S)} \wedge \Delta_2^{(S)}\right) \vee 1}$

in ℓ^1-type norm. The proof consists in approximating the signal F_{SS} by a circulant and circular-R matrix (Definition D.8) whose spectrum is known (Lemma D.9) and provides information on the latent positions $\mathbf{x}_S^*$. The difference between the spectrums of F_{SS} and A_{SS} will be bounded using the Davis-Kahan perturbation bound.

D.4.1. Preliminaries: General facts on R-matrices

Let us start by introducing the notion of circulant matrix (see [23, 33]).

Definition D.8. *For any integer $n \geq 1$, a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is circulant if there exists a vector $\mathbf{a}$ of size n such that $M_{ij} = a_{|i-j|}$ and*

$$\forall k = 1, \dots, n-1, \qquad a_k = a_{n-k}.$$

Moreover, M is a circulant and circular R-matrix if the above holds and the sequence $(a_j)_{0\leq j\leq \lfloor n/2\rfloor}$ is non-increasing.

The spectrum of circulant matrices is known –see [23] and the references therein, which allows to easily deduce the spectrum of symmetric circulant matrices, see Proposition C.4 from [33]. For clarity, we recall this result below –with a small correction on the first coordinate of the eigenvector $\mathbf{v}^{(m)}$.

Lemma D.9 (spectrum of symmetric circulant matrices)**.** *Let $M \in \mathbb{R}^{n\times n}$ be any symmetric circulant matrix associated to a vector $\mathbf{a}$ (as above).*

- *For $n = 2p+1$, the eigenvalues of M are equal to*

  $$\alpha_m = a_0 + 2\sum_{j=1}^{p} a_j \cos\left(j\frac{2\pi m}{n}\right) \quad , \qquad m = 0, \dots, p \ ,$$

  where each α_m, for $m = 1, \dots, p$, has multiplicity 2 and is associated with the two following eigenvectors

  $$\begin{aligned} \mathbf{u}^{(m)} &= (1, \cos(2\pi m/n), \dots, \cos((n-1)2\pi m/n)) \\ \mathbf{v}^{(m)} &= (0, \sin(2\pi m/n), \dots, \sin((n-1)2\pi m/n)) \ . \end{aligned} \tag{67}$$

  For $m = 0$, α_0 has multiplicity 1 and is associated to $\mathbf{u}^{(0)} = (1, \dots, 1)$.
- *For $n = 2p$,*

  $$\alpha_m = a_0 + 2\sum_{j=1}^{p-1} a_j \cos(j\frac{2\pi m}{n}) + a_p\cos(\pi m) \ , \qquad m = 0, \dots, p \ ,$$

  where each α_m, for $m = 1, \dots, p-1$, is associated with the two eigenvectors in (67). The eigenvalue α_p is associated with $\mathbf{u}^{(p)} = (1, -1, \dots, 1, -1)$. For $m = 0$, α_0 has multiplicity 1 and is associated to $\mathbf{u}^{(0)} = (1, \dots, 1)$.

If the vector $\mathbf{a}$ has nonnegative entries, then α_0 is obviously the largest eigenvalue. The next lemma ensures that, for circular R-matrices, α_1 is the second largest eigenvalue. Its proof can be found in [33, Proposition C.5].

Lemma D.10 (second largest eigenvalue)**.** *For any symmetric and circulant circular R-matrix, with nonnegative entries and eigenvalues* $\{\alpha_m\}$ *for* $m = 0, \ldots, \lfloor n/2 \rfloor$ *(as defined in Lemma D.9), we have* $\alpha_1 \geq \alpha_j$ *for all* $j = 2, \ldots, \lfloor n/2 \rfloor$.

Remark. *If* $a_j = g(j\frac{2\pi}{n})$*, then the discrete Fourier transform* $\mathcal{F}_{k,n}(g)$ *as defined in* (25) *satisfies* $\mathcal{F}_{k,n}(g) = \alpha_k$*, for all for* $k = 0, \ldots, p$*. In addition,* $\mathcal{F}_{n-k,n}(g) = \mathcal{F}_{k,n}(g)$ *for all* $k = 1, \ldots, p$.

D.4.2. Main Proof of Proposition D.1

Recall that $S \subset \{1, 2, \ldots, n\}$ satisfies $|S| = n_0 = n/4$. For the ease of exposition we assume that $S = \{1, 2, \ldots, n_0\}$ and we only consider the case where n_0 is odd (the case of even n_0 being similar). Thus, we write $n_0 = 2p + 1$ in the following. If $\Delta_1^{(S)} \wedge \Delta_2^{(S)} \leq C_{lLea}\sqrt{n\log(n)}$, then the bound in Proposition D.1 trivially holds

$$\min_{Q\in\mathcal{O}} \|\hat{\mathbf{x}}_S^{\mathrm{VSA}} - Q\mathbf{x}_S^*\|_1 \leq 2n_0 \leq C_{lLea}\frac{n\sqrt{n\log(n)}}{\left(\Delta_1^{(S)} \wedge \Delta_2^{(S)}\right) \vee 1} \quad .$$

We assume therefore that $\Delta_1^{(S)} \wedge \Delta_2^{(S)} \geq C_{lLea}\sqrt{n\log(n)}$ for a quantity C_{lLea} that will be set later. By definition of $\Delta_1^{(S)}$ and $\Delta_2^{(S)}$, this means that

$$|\lambda_0^{*(S)} - \lambda_1^{*(S)}| \wedge |\lambda_2^{*(S)} - \lambda_3^{*(S)}| \geq C_{lLea}\sqrt{n\log(n)}. \tag{68}$$

Let $\mathbf{u}^{(1)} = (u_1^{(1)}, \ldots, u_{n_0}^{(1)})$ and $\mathbf{v}^{(1)} = (v_1^{(1)}, \ldots, v_{n_0}^{(1)})$ denote the eigenvectors of a circular and circulant R-matrix as described in Lemma D.9. For any matrix $M = (m_{ij})$, we write $\|M\|_\infty$ its entry-wise l_∞ norm, that is $\|M\|_\infty = \max_{ij} m_{ij}$. Recall that $F_{SS} := [f(x_i^*, x_j^*)]_{i,j\in S}$.

Lemma D.11. *There exist a permutation* σ *and a circulant circular R-matrix* R *with nonnegative entries such that the following inequality holds* $\|F_{SS} - R_\sigma\|_\infty \leq C_{lLea}\sqrt{\log(n)/n}$*. Besides, the vector* $\mathbf{x}_S^{**} \in \mathbf{\Pi}_{n_0}$ *defined by* $x_i^{**} := (u_{\sigma(i)}^{(1)}, v_{\sigma(i)}^{(1)})$ *for* $i = 1, \ldots, n_0$ *satisfies*

$$\min_{Q\in\mathcal{O}} d_\infty(\mathbf{x}_S^{**}, Q\mathbf{x}_S^*) \leq C_a\sqrt{\frac{\log(n)}{n}} \quad . \tag{69}$$

Denote $\lambda_0 \geq \ldots \geq \lambda_{n_0-1}$ the eigenvalues of R. Lemmas D.9 and D.10 ensure that

$$\lambda_0 = \alpha_0 \geq \lambda_1 = \alpha_1 \geq \lambda_2 = \alpha_1 \geq \lambda_j \quad , \tag{70}$$

where $\lambda_j \in \{\alpha_2, \ldots, \alpha_{\lfloor n_0/2 \rfloor}\}$ for all $j = 3, \ldots, n_0 - 1$.

Lemma D.11 ensures that there exists a constant C''_{lLea} depending only on c_l, c_L, c_e, c_a such that $\|F_{SS} - R_\sigma\|_2 \leq C''_{lLea}\sqrt{n\log(n)}$. Since R_σ has the same eigenvalues as R, it follows from Weyl's inequality (see e.g. [39, page 45]) that

$$|\lambda_i^{*(S)} - \lambda_i| \leq \|F_{SS} - R_\sigma\|_{op} \leq \|F_{SS} - R_\sigma\|_2 \leq C''_{lLea}\sqrt{n\log(n)} \ , \tag{71}$$

for all $i = 0, \ldots, n_0 - 1$.

If the constant C_{lLea} in (68) is chosen as $4C''_{lLea}$ where C''_{lLea} is introduced in (71), it follows that

$$\begin{aligned} \lambda_0 - \lambda_1 &\geq (\lambda_0^{*(S)} - \lambda_1^{*(S)}) - (\lambda_0^{*(S)} - \lambda_0) - (\lambda_1 - \lambda_1^{*(S)}) \\ &\geq (C_{lLea} - 2C''_{lLea})\sqrt{n\log(n)} \\ &\geq \frac{C_{lLea}}{2}\sqrt{n\log(n)} \ , \end{aligned} \tag{72}$$

and similarly,

$$\lambda_2 - \lambda_3 \geq (\lambda_2^{*(S)} - \lambda_3^{*(S)}) - (\lambda_2^{*(S)} - \lambda_2) - (\lambda_3 - \lambda_3^{*(S)}) \geq \frac{C_{lLea}}{2}\sqrt{n\log(n)} \ .$$

Since the eigenvectors $(\sqrt{2/n_0})\mathbf{u}^{(1)}$ and $(\sqrt{2/n_0})\mathbf{v}^{(1)}$ of R are orthonormal (see Lemma D.12 below), the vectors $(\sqrt{2/n_0})\mathbf{u}_\sigma^{(1)}$ and $(\sqrt{2/n_0})\mathbf{v}_\sigma^{(1)}$ are orthonormal eigenvectors of R_σ, with the same eigenvalue $\lambda_1 = \lambda_2 = \alpha_1$.

Lemma D.12. *The vectors* $(\sqrt{2/n_0})\mathbf{u}^{(1)}$ *and* $(\sqrt{2/n_0})\mathbf{v}^{(1)}$ *are orthonormal.*

Next, we state a variant of Davis-Kahan perturbation bound [42, see Theorem 2].

Lemma D.13 (Davis-Kahan)**.** *Let* $M, \hat{M} \in \mathbb{R}^{n_0\times n_0}$ *be two symmetric matrices, with eigenvalues* $\lambda_0 \geq \ldots \geq \lambda_{n_0-1}$ *and* $\hat{\lambda}_0 \geq \ldots \geq \hat{\lambda}_{n_0-1}$ *respectively. Fix* $0 \leq r \leq s \leq n_0 - 1$ *and assume that* $(\lambda_{r-1} - \lambda_r) \wedge (\lambda_s - \lambda_{s+1}) > 0$*, where* $\lambda_{-1} = \infty$ *and* $\lambda_n = -\infty$*. Let* $d = s - r + 1$*, and let* $\mathbf{V} = (\mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_s) \in \mathbb{R}^{n_0\times d}$ *and* $\hat{\mathbf{V}} = (\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_{r+1}, \ldots, \hat{\mathbf{v}}_s) \in \mathbb{R}^{n_0\times d}$ *have orthonormal columns satisfying* $M\mathbf{v}_j = \lambda_j\mathbf{v}_j$ *and* $\hat{M}\hat{\mathbf{v}}_j = \hat{\lambda}_j\hat{\mathbf{v}}_j$ *for* $j = r, r+1, \ldots, s$*. Then, there exists an orthogonal matrix* $Q \in \mathbb{R}^{d\times d}$ *such that*

$$\|\hat{\mathbf{V}}Q - \mathbf{V}\|_2 \leq \sqrt{8d}\frac{\|\hat{M} - M\|_{op}}{(\lambda_{r-1} - \lambda_r) \wedge (\lambda_s - \lambda_{s+1})} \ .$$

The assumptions of Lemma D.13 are therefore fulfilled for the orthonormal eigenvectors $(\sqrt{2/n_0})\mathbf{u}_\sigma^{(1)}$ and $(\sqrt{2/n_0})\mathbf{v}_\sigma^{(1)}$ and the positive spectral gaps $(\lambda_0 - \lambda_1) \wedge (\lambda_2 - \lambda_3) > 0$. Hence, for $\hat{\mathbf{x}}_S^{\mathrm{VSA}}$ and $\mathbf{x}_S^{**} = (\mathbf{u}_\sigma^{(1)}, \mathbf{v}_\sigma^{(1)})^T$ in $\mathbb{R}^{2\times n}$, Lemma D.13 entails

$$\sqrt{\frac{2}{n_0}}\|Q\hat{\mathbf{x}}_S^{\mathrm{VSA}} - \mathbf{x}_S^{**}\|_2 \lesssim \frac{\|A_{S,S} - R_\sigma\|_{op}}{(\lambda_0 - \lambda_1) \wedge (\lambda_2 - \lambda_3)}$$

for some $Q \in \mathcal{O}$.

It remains to control $\|A_{S,S} - R_\sigma\|_{op}$ and the spectral gap. Since $A_{ii} = 0$ for all i (by convention), we introduce the matrix $F^{(0)}_{SS}$ such that $F^{(0)}_{ii} = 0$ for all i, and $F^{(0)}_{ij} = F_{ij}$ for all $i \neq j$.

$$\|A_{SS} - R_\sigma\|_{op} \leq \|A_{SS} - F^{(0)}_{SS}\|_{op} + \|F^{(0)}_{SS} - F_{SS}\|_{op} + \|F_{SS} - R_\sigma\|_2 \ , \tag{73}$$

using the triangular inequality and the fact that the operator norm is smaller than the Frobenius norm. The second term $\|F^{(0)}_{SS} - F_{SS}\|_{op}$ is smaller than 1 since $F_{SS} - F^{(0)}_{SS}$ is the diagonal matrix with diagonal coefficients $f(x^*_i, x^*_i) \leq 1$, the last inequality coming from $f \in \mathcal{BL}[c_l, c_L, c_e]$. To control the operator norm of the noise matrix, we shall use the following result [40, Corollary 4.4.8]. See the same reference for the definition of sub-Gaussian norms $\|.\|_{\psi_2}$.

Lemma D.14 (norm of symmetric matrices with sub-gaussian entries)**.** *Let A be an $n_0 \times n_0$ symmetric random matrix whose entries A_{ij} on and above the diagonal are independent mean-zero sub-gaussian random variables. Then, for any $t > 0$, we have*

$$\|A\|_{op} \leq CK(\sqrt{n} + t)$$

with probability at least $1 - 4e^{-t^2}$. Here $K = \max_{i,j} \|A_{i,j}\|_{\psi_2}$.

Applying the above lemma with $t = C\sqrt{\log(n)}$ (for a large enough numerical constant C) to the difference $A_{SS} - F^{(0)}_{SS}$, we obtain $\|A_{SS} - F^{(0)}_{SS}\|_{op} \lesssim \sqrt{n}$ with probability higher than $1 - 1/n^2$.

Together with Lemma D.11 and the bound (73), we deduce that $\|A_{SS} - R_\sigma\|_{op} \leq C_{lLea}\sqrt{n\log(n)}$, so that

$$\|Q\hat{\mathbf{x}}^{\mathrm{VSA}}_S - \mathbf{x}^{**}_S\|_2 \leq C_{lLea} \frac{n\sqrt{\log(n)}}{(\lambda_0 - \lambda_1) \wedge (\lambda_2 - \lambda_3)} \ .$$

Then, we deduce from the Cauchy-Schwarz inequality that

$$\min_{Q' \in \mathcal{O}} \|\hat{\mathbf{x}}^{\mathrm{VSA}}_S - Q'\mathbf{x}^{**}_S\|_1 \leq \|Q\hat{\mathbf{x}}^{\mathrm{VSA}}_S - \mathbf{x}^{**}_S\|_1 \leq C_{lLea} \frac{n\sqrt{n\log(n)}}{(\lambda_0 - \lambda_1) \wedge (\lambda_2 - \lambda_3)} \ .$$

The bounds (68) and (71) for $C_{lLea} = 4C''_{lLea}$ allow us to replace the above spectral gaps by $(\lambda^{*(S)}_0 - \lambda^{*(S)}_1) \wedge (\lambda^{*(S)}_2 - \lambda^{*(S)}_3)$. Finally, by (69) and the equivalence between the distance d in $\mathcal{C}$ and the ℓ^1-norm in $\mathbb{R}^2$, we have $\min_{Q \in \mathcal{O}} \|\mathbf{x}^*_S - Q\mathbf{x}^{**}_S\|_1 \leq C_a\sqrt{n\log(n)}$. Since $\lambda_0 \leq n$ (all the entries of F_{SS} belong to $[0, 1]$), we then deduce from the triangular inequality that

$$\min_{Q \in \mathcal{O}} \|\hat{\mathbf{x}}^{\mathrm{VSA}}_S - Q\mathbf{x}^*_S\|_1 \leq C'_{lLea} \frac{n\sqrt{n\log(n)}}{(\lambda^{*(S)}_0 - \lambda^{*(S)}_1) \wedge (\lambda^{*(S)}_2 - \lambda^{*(S)}_3)} \ .$$

Proposition D.1 is proved. ☐

D.4.3. Proofs of technical lemmas

Proof of Lemma D.11. In this proof, we replace the notation F_{SS} by $F_{(\mathbf{x}_S^*,f)}$ for clarity. Since the vector $\mathbf{x}_S^*$ satisfies (57), there exists $\mathbf{x}_S \in \mathbf{\Pi}_{n_0}$ such that the following inequality holds

$$d_\infty(\mathbf{x}_S^*, \mathbf{x}_S) \leq C_a\sqrt{\log(n)/n} \ . \tag{74}$$

Combining this with the bi-Lipschitz condition (5), we get $\max_{ij\in[n_0]} |f(x_i^*, x_j^*) - f(x_i, x_j)| \leq C_{Lea}\sqrt{\log(n)/n}$, that is,

$$\|F_{(\mathbf{x}_S^*,f)} - F_{(\mathbf{x}_S,f)}\|_\infty \leq C_{Lea}\sqrt{\frac{\log(n)}{n}} \ , \tag{75}$$

for the matrices $F_{(\mathbf{x}_S^*,f)} := [f(x_i^*, x_j^*)]_{i,j\in[n_0]}$ and $F_{(\mathbf{x}_S,f)} := [f(x_i, x_j)]_{i,j\in[n_0]}$.

Recall that $S = \{1, \dots, n_0\}$ for the ease of exposition. Let τ be some permutation that orders $\mathbf{x}_S = (x_1, \dots, x_{n_0})$ on the unit sphere, that is, such that $x_{\tau(1)}, \dots, x_{\tau(n_0)}$ is ordered. Then, one can observe that the matrix $F_{(\mathbf{x}_S,f),\tau}$ is symmetric circulant since f is a symmetric function which satisfies the geometric condition (23) with respect to the geodesic distance d on the unit sphere $\mathcal{C}$.

The matrix $F_{(\mathbf{x}_S,f),\tau}$ is therefore defined by a single vector $\mathbf{a}$ of size n_0 as in Definition D.8 of circulant matrices. This vector satisfies $a_s = g(2\pi s/n_0)$ for $s = 0, \dots, \lfloor n_0/2 \rfloor$, where we recall that $g(d(x,y)) = f(x,y)$ in the geometric setting. From the Lipschitz condition (6), we deduce that $\mathbf{a}$ satisfies some kind of weak non-increasing condition, that is $a_t \geq a_s \geq 0$ for all $0 \leq t < s \leq \lfloor n_0/2 \rfloor$ such that $s - t \geq C_{le}\sqrt{n\log(n)}$.

From the bi-Lipschitz condition (5), it is easy to see that $\mathbf{a}$ can be uniformly approximated by a non-increasing vector $\mathbf{a}'$ such that $\max_j |a_j - a_j'| \leq C_{lLe}\sqrt{\log(n)/n}$. Denoting R the circulant circular R-matrix based on the vector $\mathbf{a}'$, this means that $\max_{ij} |R_{ij} - f(x_{\tau(i)}, x_{\tau(j)})| \leq C_{lLe}\sqrt{\log(n)/n}$. Hence,

$$\|F_{(\mathbf{x}_S,f)} - R_{\tau^{-1}}\|_\infty = \|F_{(\mathbf{x}_S,f),\tau} - R\|_\infty \leq C_{lLe}\sqrt{\frac{\log(n)}{n}} \ .$$

The first result of Lemma D.11 is a consequence of (75) and the last display, setting $\sigma = \tau^{-1}$.

Next, by definition of τ, the vector $(\mathbf{x}_S)_{\sigma^{-1}} = (\mathbf{x}_S)_\tau$ is ordered, and it therefore equals any other ordered vector in $\mathbf{\Pi}_{n_0}$ up to an orthogonal transformation. Hence, we have $[(\mathbf{x}_S)_{\sigma^{-1}}]_i = Q(u_i^{(1)}, v_i^{(1)})$ for some orthogonal transformation Q in $\mathbb{R}^2$, by definition of $\mathbf{u}^{(1)}$ and $\mathbf{v}^{(1)}$. Equivalently, we have $(\mathbf{x}_S)_i = Q((\mathbf{u}_\sigma^{(1)})_i, (\mathbf{v}_\sigma^{(1)})_i)$. Then, we conclude again from (74) that the vector $\mathbf{x}_S^{**} := (\mathbf{u}_\sigma^{(1)}, \mathbf{v}_\sigma^{(1)})$ satisfies the second result of the lemma. □

Proof of Lemma D.12. Since $\sum_{k=0}^{n_0-1} e^{\iota 4\pi k/n_0} = 0$, we have $\sum_{k=0}^{n_0-1} \cos(4\pi k/n_0) = 0$ and $\sum_{k=0}^{n_0-1} \sin(4\pi k/n_0) = 0$. Then, combining with the trigonometric formulas, $\cos(2x) = 2\cos^2(x) - 1$, and $\sin(2x) = 2\cos(x)\sin(x)$, we get

$$\|\mathbf{u}^{(1)}\|_2^2 = \sum_{k=0}^{n_0-1} \cos^2(2\pi k/n_0) = \frac{n_0}{2} \ , \tag{76}$$

$$\langle \mathbf{u}^{(1)}, \mathbf{v}^{(1)} \rangle = \sum_{k=0}^{n_0-1} \cos(2\pi k/n_0)\sin(2\pi k/n_0) = 0 \ .$$

Besides, $\|\mathbf{u}^{(1)}\|_2^2 + \|\mathbf{v}^{(1)}\|_2^2 = n_0$ since $(u_i^{(1)}, v_i^{(1)})$ for any $i \in [n_0]$ is a point of the unit sphere $\mathcal{C}$. The combination with (76) leads to $\|\mathbf{v}^{(1)}\|_2^2 = n_0/2$. ☐

D.5. Proof of Lemma 4.2

Similarly to the proof of Lemma D.11, we consider a vector $\mathbf{x} \in \mathbf{\Pi}_n$ achieving $d_\infty(\mathbf{x}^*, \mathbf{x}) = d_\infty(\mathbf{x}^*, \mathbf{\Pi}_n) \leq C_a\sqrt{\log(n)/n}$. By the Bi-Lipschitz condition, the matrices $F_{\mathbf{x}^*,f}$ and $F_{\mathbf{x},f}$ satisfy

$$\|F_{\mathbf{x},f} - F_{\mathbf{x}^*,f}\|_\infty \leq C_{Lea}\sqrt{\frac{\log(n)}{n}} \ .$$

Since f is geometric and $\mathbf{x}$ belongs to $\mathbf{\Pi}_n$, it follows that, up to a permutation, $F_{\mathbf{x},f}$ is a symmetric circulant matrix associated to the vector $a_j = g(j\frac{2\pi}{n})$ for $j = 0, \dots, \lfloor \frac{n}{2} \rfloor$. It then follows from Lemma D.9, that the eigenvalues of $F_{\mathbf{x},f}$ are equal to the discrete Fourier transform $\mathcal{F}_{m,n}(g)$ of g.

The sequence (a_j), for $j = 0, \dots, \lfloor \frac{n}{2} \rfloor$ is not non-increasing because the function g is not exactly decreasing with respect to the distance. Still, arguing as in the proof of Lemma D.11, we can build an non-increasing sequence $\mathbf{a}'$ satisfying $\max_j |a_j - a'_j| \leq C_{lLe}\sqrt{\log(n)/n}$. The eigenvalues of the corresponding circulant and circular R-matrix R are also given by Lemma D.9. We denote them $\alpha_0, \alpha_1, \dots, \alpha_{\lfloor n/2 \rfloor}$. It follows from the definition of the Fourier transform that

$$|\mathcal{F}_{m,n}(g) - \alpha_m| \leq C_{lLea}\sqrt{n\log(n)} \ , \qquad 0 \leq m \leq \lfloor \frac{n}{2} \rfloor \ .$$

Hence, the gaps in the Fourier transform $\Phi_1 = \mathcal{F}_{0,n}(g) - \mathcal{F}_{1,n}(g)$ and $\Phi_2 = \min_{j=2,\dots,\lfloor n/2 \rfloor} \ \mathcal{F}_{1,n}(g) - \mathcal{F}_{j,n}(g)$ satisfy

$$\Big|\Phi_1 - (\alpha_0 - \alpha_1)\Big| \vee \Big|\Phi_2 - \min_{m=2,\dots,\lfloor \frac{n}{2} \rfloor} (\alpha_1 - \alpha_m)\Big| \leq C_{lLea}\sqrt{n\log(n)} \ .$$

To conclude, it remains to prove that

$$\Big|(\alpha_0 - \alpha_1) - \Delta_1\Big| \leq C'_{lLea}\sqrt{n\log(n)} \ ; \tag{77}$$

$$\Big|\min_{m=2,\dots,\lfloor \frac{n}{2} \rfloor} (\alpha_1 - \alpha_m) - \Delta_2\Big| \leq C'_{lLea}\sqrt{n\log(n)} \ .$$

By Lemma D.10, we have $\alpha_0 > \alpha_1 \geq \max_{j=2,\ldots,\lfloor \frac{n}{2} \rfloor} \alpha_j$. Hence, if we denote $\lambda_0 \geq \lambda_1 \ldots \geq \lambda_{n-1}$ the ordered eigenvalues of R we have

$$\lambda_0 = \alpha_0 \geq \lambda_1 = \alpha_1 \geq \lambda_2 = \alpha_1 \geq \lambda_3 \ , \tag{78}$$

where $\lambda_3 = \max\{\alpha_2, \ldots, \alpha_{\lfloor n/2 \rfloor}\}$. By definition of $\mathbf{a}'$ and $\mathbf{x}$, there exists a permutation τ of $[n]$ such that

$$\|F_{\mathbf{x}^*,f} - R_\tau\|_\infty \leq C_{lLea}\sqrt{\frac{\log(n)}{n}} \ ,$$

which implies that $\|F_{\mathbf{x}^*,f} - R_\tau\|_{op} \leq \|F_{\mathbf{x}^*,f} - R_\tau\|_2 \leq C_{lLea}\sqrt{n\log(n)}$. Denoting $\lambda_0^* \geq \lambda_1^* \geq \lambda_2^* \geq \ldots$ the ordered eigenvalues of $F_{\mathbf{x}^*,f}$, we deduce from Weyl's inequality that

$$\max_{j=0,\ldots,n-1} |\lambda_j^* - \lambda_j| \leq C_{lLea}\sqrt{n\log(n)} \ .$$

Together with (78), we deduce that $\Delta_1 = \lambda_0^* - \lambda_1^*$ and $\Delta_2 = \lambda_2^* - \lambda_3^*$ satisfy (77) which concludes the proof.

D.6. Proof of Corollary 4.3 (spectral gap for affine functions)

We will show that, for n large enough, the gaps in the Fourier $\Phi_1 := \mathcal{F}_{0,n}(g) - \mathcal{F}_{1,n}(g)$ and $\Phi_2 := \min_{j=2,\ldots,\lfloor n/2 \rfloor} \ \mathcal{F}_{1,n}(g) - \mathcal{F}_{j,n}(g)$ are at least of the the order of n. Corollary 4.3 will then follow directly from Theorem 4.1 and Lemma 4.2.

Recall that the m-th coefficient Fourier transform is defined as

$$\mathcal{F}_{m,n}(g) = \sum_{j=0}^{n-1} g\left(j\frac{2\pi}{n}\right)\cos\left(j\frac{2\pi m}{n}\right) \ .$$

For simplicity, we only consider the case where n is odd – the case of even n being similar. Let $n = 2p+1$ with $p \geq 2$. Using the fact that $g(x) = 1 - x/(2\pi)$, we get

$$\mathcal{F}_{m,n}(g) = 1 + 2\sum_{j=1}^{p}(1 - \frac{j}{n})\cos\left(j\frac{2\pi m}{n}\right) \ , \qquad m = 0, \ldots, p \ . \tag{79}$$

For convenience, $\mathcal{F}_{m,n}(g)$ is denoted by α_m in the sequel. Let us show that, for n large enough, $(\alpha_0 - \alpha_1) \gtrsim n$ and $\min_{m\geq 2}(\alpha_1 - \alpha_m) \gtrsim n$. For $m = 0$,

$$\alpha_0 = 1 + 2p - \frac{2}{n}\sum_{j=1}^{p} j = 1 + 2p - \frac{p(p+1)}{n} = n - \frac{(n-1)(n+1)}{4n} = \frac{3n}{4} + \frac{1}{4n} \ . \tag{80}$$

For $m \geq 1$, we can still work out explicitly α_m.

$$\alpha_m = -2\sum_{j=0}^{p}\frac{j}{n}\cos\left(j\frac{2\pi m}{n}\right) = -\frac{2}{n}Re\left[\sum_{j=1}^{p} je^{\iota j 2\pi m/n}\right] \tag{81}$$

$$= -\frac{2}{n} Re\left[-\iota f'\left(\frac{2\pi m}{n}\right)\right] \quad ,$$

where the function f is defined as $f(x) := \sum_{j=0}^{p} e^{\iota j x} = \frac{e^{\iota(p+1)x}-1}{e^{\iota x}-1}$ for $x \in (0, \pi)$. We work out $f'(x)$:

$$\begin{aligned} -\iota f'(x) &= \frac{(p+1)e^{\iota(p+1)x}(e^{\iota x}-1) - e^{\iota x}(e^{\iota(p+1)x}-1)}{(e^{\iota x}-1)^2} \\ &= -\frac{(p+1)e^{\iota p x}(e^{\iota x}-1) - (e^{\iota(p+1)x}-1)}{4\sin^2(x/2)}, \\ &= -\iota \frac{(p+1)e^{\iota(p+\frac{1}{2})x}\sin(\frac{x}{2}) - e^{\iota\frac{p+1}{2}x}\sin(\frac{p+1}{2}x)}{2\sin^2(x/2)}, \end{aligned}$$

where the second line follows from $(e^{\iota x}-1)^2 = -4\sin^2(x/2)e^{\iota x}$. Hence,

$$Re\left[-\iota f'(x)\right] = \frac{(p+1)\sin((p+\frac{1}{2})x)\sin(\frac{x}{2}) - \sin^2(\frac{p+1}{2}x)}{2\sin^2(x/2)} \quad .$$

Taking $x_m = \frac{2\pi m}{n}$, the first term of the numerator is equal to zero since $\sin((p+\frac{1}{2})x_m) = \sin(\pi m) = 0$. Then, combining the above with (81) yields

$$\alpha_m = -\frac{2}{n} Re\left[-\iota f'\left(\frac{2\pi m}{n}\right)\right] = \frac{\sin^2(\frac{p+1}{2}x_m)}{n\sin^2(\frac{\pi m}{n})} \quad .$$

Since $\frac{p+1}{2}x_m = m\frac{\pi}{2} + m\frac{\pi}{2n}$,

$$\alpha_m = \begin{cases} \frac{\cos^2(\frac{m\pi}{2n})}{n\sin^2(\frac{\pi m}{n})} & \text{if } m \text{ is odd,} \\ \\ \frac{\sin^2(\frac{m\pi}{2n})}{n\sin^2(\frac{\pi m}{n})} & \text{if } m \text{ is even.} \end{cases}$$

Hence, the sequence of eigenvalues with odd indices is decreasing:

$$\alpha_1 > \alpha_3 > \alpha_5 > \ldots > \ldots \quad ,$$

since the fraction $\cos^2(m\pi/2n)/\sin^2(\pi m/n)$ decreases with $m \in [p]$. In the (remaining) case of even indices, the numerator can be upper bounded as follows: $\sin^2(\frac{m\pi}{2n}) \leq \sin^2(\pi/4) \leq \min_{k\in[p]}\cos^2(\frac{k\pi}{2n})$, which leads to $\alpha_{2r} \leq \alpha_{2r-1}$ for all $r = 1, \ldots, \lfloor p/2 \rfloor$.

In other words, each eigenvalue of even index is upper bounded by the previous eigenvalue. In light of this, we only need to prove that, for n large enough,

$$(\alpha_0 - \alpha_1) \wedge (\alpha_1 - \alpha_2) \wedge (\alpha_1 - \alpha_3) \gtrsim n \quad .$$

From (80), we deduce that α_0 is equivalent to $3n/4$. Besides, we deduce from the explicit form of α_m in the general case that α_1, α_2, and α_3 are respectively equivalent to $\frac{n}{\pi^2}$, $1/(4n)$, and $n/(4\pi^2)$. This completes the proof. □

D.7. Proof of Proposition 4.4

For the first inequality of the proposition, the proof is the same as for Proposition D.1, after replacing $\hat{\mathbf{x}}_S^{\mathrm{VSA}}, \mathbf{x}_S^*$, A_{SS} respectively by $\hat{\mathbf{x}}^{\mathrm{VSA}}, \mathbf{x}^*$, A. The second inequality of the proposition follows from Lemma 4.2. $\square$

Appendix E: Probabilistic inequalities

We recall Hoeffding inequality for hypergeometric distributions.

Lemma E.1. *For $N \geq 1$, $p \in [0,1]$ and $n \geq N$, let X be a hypergeometric random variable with parameters (N, p, n). Then, for all $t > 0$,*

$$\mathbb{P}\left(|X - Np| \geq \sqrt{\frac{Nt}{2}}\right) \leq 2e^{-t} \quad . \tag{82}$$

References

[1] Arias-Castro, E., Javanmard, A. and Pelletier, B. (2020). Perturbation bounds for procrustes, classical scaling, and trilateration, with applications to manifold learning. *Journal of Machine Learning Research* **21** 15–1. MR4071198

[2] Athreya, A., Tang, M., Park, Y. and Priebe, C. E. (2021). On estimation and inference in latent structure random graphs. *Statistical Science* **36** 68–88. MR4194204

[3] Atkins, J. E., Boman, E. G. and Hendrickson, B. (1998). A spectral algorithm for seriation and the consecutive ones problem. *SIAM Journal on Computing* **28** 297–310. MR1630473

[4] Ballinger, T. P. and Wilcox, N. T. (1997). Decisions, error and heterogeneity. *The Economic Journal* **107** 1090–1105. `https://doi.org/10.1111/j.1468-0297.1997.tb00009.x`

[5] Barnard, S. T., Pothen, A. and Simon, H. (1995). A spectral algorithm for envelope reduction of sparse matrices. *Numerical linear algebra with applications* **2** 317–334. MR1349276

[6] Bradley, R. A. and Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* **39** 324–345. MR0070925

[7] Chen, K. and Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association* **113** 241–251. MR3803461

[8] Chen, Y., Fan, J., Ma, C. and Wang, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *The Annals of Statistics* **47** 2204–2235. MR3953449

[9] Chen, Y. and Suh, C. (2015). Spectral MLE: Top-K Rank Aggregation from Pairwise Comparisons. In *Proceedings of the 32nd International Con-*

ference on Machine Learning (F. Bach and D. Blei, eds.). *Proceedings of Machine Learning Research* **37** 371–380. PMLR, Lille, France.

[10] Coifman, R. R., Shkolnisky, Y., Sigworth, F. J. and Singer, A. (2008). Graph Laplacian Tomography From Unknown Random Projections. *IEEE Transactions on Image Processing* **17** 1891–1899. MR2517273

[11] De Castro, Y., Lacour, C. and Ngoc, T. M. P. (2017). Adaptive estimation of nonparametric geometric graphs. arXiv preprint arXiv:1708.02107. MR4165265

[12] Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs. arXiv preprint arXiv:0712.2749. MR2463439

[13] Diaz, J., McDiarmid, C. and Mitsche, D. (2020). Learning random points from geometric graphs or orderings. *Random Structures & Algorithms* **57** 339–370. MR4129724

[14] Elson, J., Karp, R. M., Papadimitriou, C. H. and Shenker, S. (2004). Global Synchronization in Sensornets. In *LATIN 2004: Theoretical Informatics* (M. Farach-Colton, ed.) 609–624. Springer Berlin Heidelberg, Berlin, Heidelberg. MR2095234

[15] Fogel, F., Jenatton, R., Bach, F. and d'Aspremont, A. (2013). Convex relaxations for permutation problems. In *Advances in Neural Information Processing Systems* 1016–1024. MR3418224

[16] Fulkerson, D. and Gross, O. (1965). Incidence matrices and interval graphs. *Pacific journal of mathematics* **15** 835–855. MR0186421

[17] Gao, C., Ma, Z., Zhang, A. Y. and Zhou, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research* **18** 1980–2024. MR3687603

[18] Gao, C. and Zhang, A. Y. (2019). Iterative algorithm for discrete structure recovery. arXiv preprint arXiv:1911.01018.

[19] Gao, C. and Zhang, A. Y. (2020). Exact Minimax Estimation for Phase Synchronization. arXiv preprint arXiv:2010.04345. MR4346085

[20] Garriga, G. C., Junttila, E. and Mannila, H. (2011). Banded structure in binary matrices. *Knowledge and information systems* **28** 197–226.

[21] Gilbert, E. N. (1961). Random Plane Networks. *Journal of the Society for Industrial and Applied Mathematics* **9** 533–543. MR0132566

[22] Giridhar, A. and Kumar, P. R. (2006). Distributed Clock Synchronization over Wireless Networks: Algorithms and Analysis. In *Proceedings of the 45th IEEE Conference on Decision and Control* 4915–4920.

[23] Gray, R. M. (2006). *Toeplitz and circulant matrices: A review*. now publishers inc.

[24] Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association* **97** 1090–1098. MR1951262

[25] Janssen, J. and Smith, A. (2020). Reconstruction of Line-Embeddings of Graphons. arXiv preprint arXiv:2007.06444. MR4361745

[26] Lei, J. and Zhu, L. (2014). A generic sample splitting approach for refined community recovery in stochastic block models. arXiv preprint arXiv:1411.1469. MR3701502

[27] Lovász, L. (2012). *Large networks and graph limits* **60**. American Mathematical Soc. MR3012035

[28] Lyzinski, V., Tang, M., Athreya, A., Park, Y. and Priebe, C. E. (2017). Community Detection and Classification in Hierarchical Stochastic Blockmodels. *IEEE Transactions on Network Science and Engineering* **4** 13–26. MR3625952

[29] Negahban, S., Oh, S. and Shah, D. (2017). Rank Centrality: Ranking from Pairwise Comparisons. *Operations Research* **65** 266–287. MR3613103

[30] Penrose, M. (2003). *Random geometric graphs* **5**. Oxford university press. MR1986198

[31] Queyranne, M. (1986). Performance ratio of polynomial heuristics for triangle inequality quadratic assignment problems. *Operations Research Letters* **4** 231–234. `https://doi.org/10.1016/0167-6377(86)90007-6` MR0829886

[32] Recanati, A., Brüls, T. and d'Aspremont, A. (2017). A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. *Bioinformatics* **33** 3188–3194.

[33] Recanati, A., Kerdreux, T. and d'Aspremont, A. (2018). Reconstructing Latent Orderings by Spectral Clustering. arXiv preprint arXiv:1807.07122.

[34] Robinson, W. S. (1951). A Method for Chronologically Ordering Archaeological Deposits. *American Antiquity* **16** 293–301.

[35] Sahni, S. and Gonzalez, T. (1976). P-Complete Approximation Problems. *J. ACM* **23** 555–565. MR0408313

[36] Shah, N., Balakrishnan, S., Guntuboyina, A. and Wainwright, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20. MR3604649

[37] Singer, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis* **30** 20–36. MR2737931

[38] Sussman, D. L., Tang, M. and Priebe, C. E. (2013). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE transactions on pattern analysis and machine intelligence* **36** 48–57.

[39] Tao, T. (2012). *Topics in random matrix theory* **132**. American Mathematical Soc. MR2906465

[40] Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science* **47**. Cambridge university press. MR3837109

[41] Yu, B. (1997). Assouad, Fano, and le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer. MR1462963

[42] Yu, Y., Wang, T. and Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006

[43] Zhang, A. Y. and Zhou, H. H. (2016). Minimax rates of community detection in stochastic block models. *The Annals of Statistics* **44** 2252–2280. MR3546450