



HAL
open science

Minimax optimal seriation in polynomial time

Yann Issartel, Christophe Giraud, Nicolas Verzelen

► **To cite this version:**

Yann Issartel, Christophe Giraud, Nicolas Verzelen. Minimax optimal seriation in polynomial time. 2024. hal-04575332

HAL Id: hal-04575332

<https://hal.science/hal-04575332v1>

Preprint submitted on 14 May 2024

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.

Minimax optimal seriation in polynomial time

Yann Issartel* and Christophe Giraud† and Nicolas Verzelen‡

May 14, 2024

Abstract

We consider the statistical seriation problem, where the statistician seeks to recover a hidden ordering from a noisy observation of a permuted Robinson matrix. In this paper, we tightly characterize the minimax rate for this problem of matrix reordering when the Robinson matrix is bi-Lipschitz, and we also provide a polynomial time algorithm achieving this rate; thereby answering two open questions of [Giraud et al., 2021]. Our analysis further extends to broader classes of similarity matrices.

1 Introduction

The seriation problem consists in ordering n objects from pairwise measurements. It has its roots in archaeology, in particular for the chronological dating of graves [Robinson, 1951]. In modern data science, seriation arises in various applications, such as envelope reduction for sparse matrices [Barnard et al., 1995], reads alignment in *de novo* sequencing [Garriga et al., 2011, Recanati et al., 2017], time synchronization in distributed networks [Elson et al., 2004, Giridhar and Kumar, 2006], or interval graph identification [Fulkerson and Gross, 1965].

1.1 Seriation problem

We are given a matrix $A = [A_{ij}]_{1 \leq i, j \leq n}$ of noisy measurements of pairwise similarities between n objects. In the seriation paradigm, it is assumed that there exists an unknown ordering (i.e. a permutation) π of $[n]$ such that, the noisy similarity A_{ij} tends to be large when π_i is close to π_j , while A_{ij} tends to be small when π_i is far from π_j . This is formalized by the following model.

Model. Similarly as in [Fogel et al., 2013, Recanati et al., 2018, Janssen and Smith, 2020, Giraud et al., 2021, Natic and Smith, 2021], we assume that $A \in \mathbb{R}^{n \times n}$ is symmetric with null diagonal, and we model A as a noisy observation of an unknown permuted Robinson matrix. More precisely, a symmetric matrix $F \in \mathbb{R}^{n \times n}$ is Robinson (or a R -matrix), if all its rows and

*LTCI, Télécom Paris, Institut Polytechnique de Paris, France. Yann.Issartel@telecom-paris.fr

†Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, CNRS, France. Christophe.Giraud@universite-paris-saclay.fr

‡MISTEA, INRAE, Institut Agro, Univ. Montpellier, France. Nicolas.Verzelen@inrae.fr

all its columns are unimodal. In other words, the entries of F decrease when moving away from the diagonal. Formally, for any $i, j \in [n]$, $i < j$, we have

$$F_{ik} < F_{jk} \text{ for all } k > j, \quad \text{and} \quad F_{ik} > F_{jk} \text{ for all } k < i. \quad (1)$$

The observed similarity matrix A is thus a noisy version of a matrix $F_\pi = [F_{\pi_i \pi_j}]_{1 \leq i, j \leq n}$,

$$A = F_\pi + \sigma E, \quad (2)$$

where $\pi : [n] \rightarrow [n]$ is an unknown permutation, and F_π is a π -permuted version of an unknown Robinson matrix $F \in [0, 1]^{n \times n}$ with null diagonal. The scalar $\sigma > 0$ corresponds to the noise level, and the noise matrix $E \in \mathbb{R}^{n \times n}$ is symmetric with null diagonal, and with lower diagonal entries distributed as independent sub-Gaussian random variables [Rigollet and Hütter, 2023, Definition 1.2] with zero means $\mathbb{E} E_{ij} = 0$ and variance proxies smaller than 1. This sub-Gaussian assumption includes the special case of Gaussian noise, and also Bernoulli observations which are used to model random graphs. Results in the sub-Gaussian setting (2) will thus hold for random graph models, where the $\{A_{ij} : i < j\}$ follow independent Bernoulli distributions with parameters $F_{\pi_i \pi_j}$.

Objective. In the noisy seriation problem, we want to recover the permutation π from the matrix A in (2). In the noiseless case ($\sigma = 0$), [Atkins et al., 1998] has proved that it is possible to efficiently reconstruct π from A by a spectral procedure. However, real-world data are often noisy, and the observation A may not be exactly a disordered R-matrix, hence the importance of handling the noisy case ($\sigma > 0$) [Fogel et al., 2013]. The noise matrix σE can make it impossible to recover π exactly, and the aim is therefore to build an estimator $\hat{\pi}$ which recovers π up to some error. As usual in noisy seriation [Janssen and Smith, 2020, Giraud et al., 2021, Natick and Smith, 2021], we consider the maximum ordering error: $\max_{i \in [n]} |\hat{\pi}_i - \pi_i|/n$. Our goal is thus to simultaneously recover all the positions π_i for $i \in [n]$. Because of a minor lack of identifiability in the seriation problem, it is impossible to decipher from A if the latent permutation is π or the reverse $\pi^{rev}(\cdot) = \pi(n - \cdot + 1)$, and hence it is usual to take the minimum of ordering errors against π and π^{rev} :

$$L_{\max}(\hat{\pi}, \pi) = \left(\max_{i \in [n]} \frac{|\hat{\pi}_i - \pi_i|}{n} \right) \wedge \left(\max_{i \in [n]} \frac{|\hat{\pi}_i - \pi_i^{rev}|}{n} \right), \quad (3)$$

where we used the notation $a \wedge b = \min\{a, b\}$ for any numbers a, b . Our objective is to build an efficient estimator $\hat{\pi}$ that achieves, with high probability, a small L_{\max} error. The optimal rates will be non-asymptotic, and characterized as functions of the problem parameters: the sample size n , the noise level σ , and some regularity of F (to be defined).

1.2 Main Contribution

A first idea would be to apply the spectral algorithm of Atkins et al. [Atkins et al., 1998] to noisy data. Under strong assumptions on the matrix F , the spectral procedure turns out to work well [Fogel et al., 2013, Giraud et al., 2021]. For instance, this is the case when the matrix F is Toeplitz and exhibits a large spectral gap. Unfortunately, previous works [Rocha et al., 2018, Janssen and Smith, 2020, Giraud et al., 2021, Cai and Ma, 2022, Briend et al., 2024] suggest that the spectral procedure performs poorly beyond this specific example. Therefore,

there is a need for new seriation algorithms that perform well even if the matrix F is not Toeplitz.

The Robinson assumption alone may be insufficient to recover π within a small estimation error. As an extreme example, F may be nearly flat and buried under the noise σE , thus leaving no chance of finding π . To exclude this trivial case, [Giraud et al., 2021] strengthen the Robinson assumption by introducing a general bi-Lipschitz assumption. It constrains the decay of F 's rows and columns, so that the variations $|F_{ik} - F_{jk}|$ are bounded from above and below by $|i - j|/n$ up to some constant factors

$$\alpha \frac{|i - j|}{n} \leq |F_{ik} - F_{jk}| \leq \beta \frac{|i - j|}{n} \quad \text{for some } k. \quad (4)$$

The proper definition of *bi-Lipschitz* matrices is given in Assumption 2.1, page 5. General models like (4) offer flexibility to fit data, but they are often a roadblock to prove sharp theoretical results. [Giraud et al., 2021] prove that the optimal rate for the L_{\max} loss (3) is of the order of $\sqrt{\log(n)/n}$. Unfortunately, their result is mainly theoretical, as the time-complexity of their algorithm is super polynomial. In addition, they do not provide dependencies in the problem parameters (α, β, σ) ; for example, one would expect $L_{\max}(\hat{\pi}, \pi)$ to decrease with σ , according to the intuition that the seriation problem becomes easier when the noise level $\sigma \rightarrow 0$ goes to zero. This left open two important questions of different nature: Does there exist efficient (i.e. polynomial-time) algorithms converging at the optimal rate $\sqrt{\log(n)/n}$? Is it possible to prove sharp rates with explicit interpretable dependencies in the problem constants (α, β, σ) ? We answer to these two questions for bi-Lipschitz Robinson matrices by:

1. Proving that the optimal minimax rate for recovering π is $L_{\max}(\hat{\pi}, \pi) = O\left(\frac{\sigma}{\alpha} \sqrt{\frac{\log(n)}{n}}\right)$;
2. Providing a polynomial-time algorithm that achieves this rate.

Our new procedure, Seriation by Aggregation of Local Bisections (SALB), is based on the following general idea. To recover the permutation π , we partially reconstruct the comparison matrix $H^{(\pi)} \in \{-1, 0, 1\}^{n \times n}$ defined by $H_{ij}^{(\pi)} = \text{sign}(\pi_i - \pi_j) = 1 - 2\mathbb{1}_{\pi_i < \pi_j}$ for all $i, j \in [n]$. In order to reconstruct the $H_{ij}^{(\pi)}$'s, we will start by estimating a distance matrix D^* such that D_{ij}^* stands for a measure of dissimilarity between i and j in $[n]$. Then, we will construct from this estimator of D^* a two-step estimator of H^π . The time complexity of SALB is polynomial in the sample size n .

For the performance analysis of SALB, we first consider the assumption (4) and we prove that, up to a numerical factor, the optimal maximum error is $(\sigma/\alpha)\sqrt{\log(n)/n}$. The ratio (σ/α) between the noise level σ and the minimal slope α of the bi-Lipschitz matrix F gives a simple and explicit dependency in the problem parameters. In a second part of the paper, we extend the analysis of SALB to more general matrices, namely (i) matrices F with only average constraints on their columns, and (ii) latent bi-Lipschitz matrices which are direct generalizations of bi-Lipschitz matrices to latent space models. We will see that the general matrices (i-ii) have desired features that bi-Lipschitz matrices do not possess. For example, the matrices (i) allow quasi null similarities ($F_{ij} \approx 0$) for objects i, j that are far apart from each other in the ordering π ; and the matrices (ii) allow the matrix F to have heterogeneous columns, thus offering more flexibility to fit the data in applications.

1.3 Related literature

There is a wide range of learning problems where the data is disordered by an unknown permutation; popular examples includes ranking [Braverman and Mossel, 2009, Mao et al., 2018b, Chen et al., 2019, Chatterjee and Mukherjee, 2019], feature matching [Collier and Dalalyan, 2016, Galstyan et al., 2022], matrix estimation under shape constraints [Flammarion et al., 2019, Chatterjee and Mukherjee, 2019, Mao et al., 2018a] and, closer to our paper, seriation in R-matrices [Atkins et al., 1998, Fogel et al., 2013, Fogel et al., 2014, Recanati et al., 2018], [Janssen and Smith, 2020, Giraud et al., 2021, Cai and Ma, 2022]. However, each problem has its own setting and goal, and hence the solutions are not always related.

Contrasting with the aforementioned literature on matrix estimation, permutation recovery in seriation has received little attention from statisticians. In particular, most existing works in the seriation problem have focused on the noiseless case ($\sigma = 0$). Efficient algorithms have been proposed using spectral methods [Atkins et al., 1998] and convex optimization [Fogel et al., 2013]. Exact recovery have been proved for R-matrices [Atkins et al., 1998] and toroidal R-matrices [Recanati et al., 2018] using spectral algorithms. As said earlier, such spectral algorithms are unfortunately not robust to noise perturbations.

Noisy seriation has recently gained interest [Janssen and Smith, 2020, Giraud et al., 2021, Natick and Smith, 2021, Cai and Ma, 2022]. Efficient algorithms are analyzed inside the set of Toeplitz R-matrices (a.k.a monotone Toeplitz matrices, see Example 1 in appendix A), which are matrices defined from a single vector. Some of the aforementioned papers may focus on a latent space model instead of a matrix model; in this case Toeplitz R-matrices are equivalent to a geometric latent space model (Example 3 in appendix A). Spectral algorithms [Giraud et al., 2021, Natick and Smith, 2021] leverage the special connection between the spectrum of Toeplitz matrices and the latent ordering π ; see [Recanati et al., 2018] for a short presentation of this connection. The iterative sorting algorithm [Cai and Ma, 2022] exploits the correlation between the position π_i and the score $S_i = \sum_{\ell} A_{i\ell}$. Unfortunately, both arguments (the spectral property and the score correlation) seem to have a limited scope: they are useful inside the set of Toeplitz matrices, but there is no clear reason why they will still work beyond these matrices. In the particular case of network data $A \in \{0, 1\}^{n \times n}$, the authors in [Janssen and Smith, 2020] study the popular graphon model and propose an algorithm based on a thresholded version of the squared adjacency matrix $A^{(2)}$. Their assumptions include a Toeplitz matrix (called “uniformly embedded graphon”) and a technical condition on the square of the graphon, to ensure that the thresholded squared adjacency matrix used in their algorithm is close to a Robinson matrix.

By contrast, we will not assume that the matrix F is Toeplitz but is rather bi-Lipschitz to analyze the performance of our efficient procedure – SALB. As discussed earlier, the bi-Lipschitz assumption has already been considered in [Giraud et al., 2021], but for this (whole) class of matrices the authors only provide a non-efficient algorithm. Also, our objective of controlling the maximum error is the same as in [Janssen and Smith, 2020, Giraud et al., 2021, Natick and Smith, 2021] but is fundamentally different from that of exact matrix reordering in [Cai and Ma, 2022]. For a more detailed comparison with [Giraud et al., 2021, Cai and Ma, 2022], we refer to the discussion below Theorem 2.1. For comparisons with results in latent space settings [Janssen and Smith, 2020, Giraud et al., 2021, Natick and Smith, 2021], see the discussion below Theorem 5.1. We emphasize that seriation in matrix settings and in latent space models are closely related. In particular, Theorem 2.1 in the matrix setting is just a

particular case of the Theorem 5.1 in the latent space setting.

Organization. The main results for bi-Lipschitz matrices are stated in Section 2, whereas our polynomial-time procedure SALB is described in Section 3. Sections 4 and 5 are dedicated to extensions to more general models. A conclusion is in section 6. Proofs are postponed to the appendix.

Notation: We write $[n]$ for the set $\{1, \dots, n\}$. Given a set G , $\#G$ stands for its cardinality. For short, we write $\{k : \text{property P}(k) \text{ holds}\}$ for $\{k \in [n] : \text{property P}(k) \text{ holds}\}$. We write $G \setminus G'$ the set $\{k : k \in G \text{ and } k \notin G'\}$. Given an $n \times n$ matrix F and a permutation $\pi : [n] \rightarrow [n]$, the permuted matrix F_π has coefficients $F_{\pi_i \pi_j}$, $i, j \in [n]$. We write $\|F_j\|_2^2 = \sum_i F_{ij}^2$ the square of the l_2 -norm of the j -th column F_j . The maximum of these vector norms is denoted by $\|F\|_{2,\infty} = \max_{j \in [n]} \|F_j\|_2$. We write $a \vee b$ for $\max(a, b)$ and $a \wedge b$ for $\min(a, b)$. The notation $a \asymp b$ means that there exist positive constants c and C such that $cb \leq a \leq Cb$. If the constants c and C depend on some parameters α, β , we use the symbol $\asymp_{\alpha, \beta}$ instead.

2 Result for bi-Lipschitz matrices

2.1 Seriation error

In this section, we focus on the subclass of Robinson matrices (1) that satisfy a bi-Lipschitz type condition.

Assumption 2.1 (bi-Lipschitz matrix). *For any constants $0 < \alpha \leq \beta$, let $\mathcal{BL}[\alpha, \beta]$ be the collection of matrices $F \in \mathbb{R}^{n \times n}$ that satisfy*

$$|F_{ik} - F_{jk}| \leq \beta \frac{|i - j|}{n} \quad \text{for all } (i, j, k) ; \quad (5)$$

$$F_{ik} - F_{jk} \geq \alpha \frac{|i - j|}{n} \quad \text{for all } k < i < j ; \quad (6)$$

$$F_{jk} - F_{ik} \geq \alpha \frac{|i - j|}{n} \quad \text{for all } i < j < k .$$

In the next theorem, we estimate the underlying permutation π in (2), using the polynomial-time algorithm SALB described in the next section. To simplify the statement of the next result, we set the tuning parameters $(\delta_1, \delta_2, \delta_3)$ in Algorithm 1 to $\delta_1 = n^{-1/5}$, $\delta_2 = \log(n)\delta_1$, and $\delta_3 = \log(n)\delta_2$.

Theorem 2.1. *For $\alpha > 0$, $\beta > 0$, $\sigma > 0$, there exists a positive constant $C_{\alpha, \beta, \sigma}$ only depending on (α, β, σ) and a numerical constant $C > 0$ such that the following holds for any $n \geq C_{\alpha, \beta, \sigma}$ and for any $F \in \mathcal{BL}[\alpha, \beta]$. With probability higher than $1 - 9/n^2$, the permutation $\hat{\pi}_o$ computed by SALB satisfies*

$$L_{\max}(\hat{\pi}_o, \pi) \leq C \frac{\sigma}{\alpha} \sqrt{\frac{\log(n)}{n}} . \quad (7)$$

In other words, the polynomial-time algorithm SALB estimates the position of each object up to an error of the order $\frac{\sigma}{\alpha} \sqrt{\frac{\log(n)}{n}}$. In particular, the error becomes smaller for larger α ,

larger n , and smaller σ . We prove in the next subsection, that the rate (7) is minimax optimal on the class $\mathcal{BL}[\alpha, \beta]$.

Theorem 2.1 above is stated in a setting where α , β , and σ are considered as constants whereas n is large. In fact, SALB achieves the convergence rate (7) even in settings where the quantities α , β , and σ depend on n , but the choices of the tuning parameters $\delta_1, \delta_2, \delta_3$ are more intricate. We refer to Theorem C.3 in appendix for a general bound.

To the best of our knowledge, $\hat{\pi}_o$ is the first estimator to both have a polynomial-time complexity and a seriation rate $\sqrt{\log(n)}/n$ over bi-Lipschitz Robinson matrices. In addition, the bound (7) captures optimal dependencies in the problem parameters (α, σ) , which is a significant improvement over [Giraud et al., 2021]. In particular, we can derive from (7) that exact recovery (i.e. $L_{\max}(\hat{\pi}_o, \pi) < 1/n$) is possible on $\mathcal{BL}[\alpha, \beta]$ as soon as $\alpha \gtrsim \sigma\sqrt{n \log(n)}$, where \gtrsim hides a numerical constant.

Exact recovery of ordering has been considered recently by [Cai and Ma, 2022] for symmetric monotone Toeplitz matrices (Example 1 in appendix A). Introducing the signal-to-noise ratio $m(\mathcal{F}) = \min_{F \in \mathcal{F}; \pi, \pi' \in \mathcal{P}} \|F_\pi - F_{\pi'}\|_F$, they prove that for some classes \mathcal{F} of (Toeplitz) signal matrices F , a minimal signal-to-noise ratio $m(\mathcal{F}) \gtrsim \sigma\sqrt{n \log(n)}$ is required for achieving exact recovery, regardless of computational considerations. The situation for signal matrices F in $\mathcal{BL}[\alpha, \beta]$ is much more favorable. Indeed, for any $0 < \alpha \leq 1$ and $\alpha \leq \beta$, considering $F_0 \in \mathcal{BL}[\alpha, \beta]$ defined by

$$(F_0)_{ij} = 1 - \alpha \frac{|i-j|}{n}, \quad \text{for all } i, j \in [n], \quad (8)$$

we observe that $m(\mathcal{BL}[\alpha, \beta]) \leq m(\{F_0\}) \leq \alpha/\sqrt{n}$; so Theorem 2.1 ensures that a signal-to-noise ratio $m(\mathcal{BL}[\alpha, \beta]) \gtrsim \sigma\sqrt{\log(n)}$ is sufficient for exact recovery in polynomial-time in $\mathcal{BL}[\alpha, \beta]$.

2.2 Optimality

In this section, we show the optimality of the rate $(\sigma/\alpha)\sqrt{\log(n)}/n$ established in (7). Given the generality of the bi-Lipschitz Assumption 2.1, one might expect that, when the data is actually drawn from a simpler parametric model, it could be possible to come up with another algorithm with faster rates of seriation. Surprisingly, this intuition turns out to be false: imposing the simpler parametric model (8) does not lead to faster rates. The following result states that, even if the statistician knows in advance that $F = F_0$, she cannot estimate π at a faster rate than (7).

In the next theorem, for any permutation π , $\mathbb{P}_{(F_0, \pi)}$ refers to the distribution of the data

$$A = (F_0)_\pi + \sigma E, \quad (9)$$

with F_0 defined by (8), and for independent Gaussian random variables $E_{ij} \sim N(0, 1)$, $i < j \in [n]$.

Theorem 2.2. *For $\alpha > 0$ and $\sigma > 0$, there exists a positive constant $C_{\alpha, \sigma}$ and a numerical positive constant c such that, for all $n \geq C_{\alpha, \sigma}$, it holds that*

$$\inf_{\hat{\pi}} \sup_{\pi} \mathbb{P}_{(F_0, \pi)} \left[L_{\max}(\hat{\pi}, \pi) \geq c \frac{\sigma}{\alpha} \sqrt{\frac{\log(n)}{n}} \right] \geq \frac{1}{2},$$

where the infimum is taken over all estimators $\hat{\pi}$, the supremum over all permutations π , and $\mathbb{P}_{(F_0, \pi)}$ is defined in (9).

Thus, for n large enough, any estimator $\hat{\pi}$ must make an error of the order at least $(\sigma/\alpha)\sqrt{\log(n)/n}$ over some permutation π , with probability at least $1/2$. The proof is provided in Appendix P.

3 Description of SALB

In this section, we describe the main ideas of our procedure Seriation by Aggregation of Local Bisections (SALB) – see Algorithm 1. For recovering the permutation π , the general idea is to partially reconstruct the comparison matrix $H^{(\pi)} \in \{-1, 0, 1\}^{n \times n}$ defined by $H_{ii}^{(\pi)} = 0$ and

$$H_{ij}^{(\pi)} = \text{sign}(\pi_i - \pi_j) = 1 - 2\mathbb{1}_{\pi_i < \pi_j} \quad \text{for all } i, j \in [n], i \neq j. \quad (10)$$

The matrix $H^{(\pi)}$ is only identifiable up to a general multiplicative factor ± 1 , but this is sufficient for our purpose. In order to reconstruct the $H_{ij}^{(\pi)}$'s, we start by estimating a similarity distance matrix D^* such that D_{ij}^* stands for a measure of dissimilarity between i and j in $[n]$ –see (11) for a definition. Algorithm 1 is mainly organized in four steps. First, we estimate the distance matrix D^* . Second, we use the distance estimate \hat{D} to perform local bisections, and by combining these local bisections, we are able to build a first estimator \hat{H}_1 of the comparison matrix $H^{(\pi)}$. Third, we build upon this estimator \hat{H}_1 to refine the estimation of $H^{(\pi)}$; we thereby obtain \hat{H}_2 . Finally, we use in Lines 4–5 a simple method to infer π from our estimator $\hat{H}_1 + \hat{H}_2$ of $H^{(\pi)}$: Since π corresponds to the permutation that ranks the objects increasingly according to the values of the row sums of $H^{(\pi)}$, we simply compute the row sums S of $\hat{H}_1 + \hat{H}_2$ and we build $\hat{\pi}_o$ by ordering the values of S . In the following subsections, we describe the three main steps (1-2-3).

Algorithm 1 Seriation by Aggregation of Local Bisections (SALB)

Require: $(A, \delta_1, \delta_2, \delta_3, \sigma)$

Ensure: $\hat{\pi}_o \in [n]^n$ an estimator of π .

- 1: $\hat{D} = \text{DistanceEstimation}(A)$ {see Algo. 4}
 - 2: $\hat{H}_1 = \text{AgregLocalBisection}(\hat{D}, \delta_1, \delta_2, \delta_3)$ {see Algo. 2}
 - 3: $\hat{H}_2 = \text{LocalRefineWS}(\hat{H}_1, \hat{D}, A, \sigma, \delta_1, \delta_2, \delta_3)$ {see Algo. 8}
 - 4: Compute the scores $S = [\hat{H}_1 + \hat{H}_2]\mathbf{1}$
 - 5: Build any permutation $\hat{\pi}_o$ by increasing values of S .
-

3.1 Distance estimation

Define the similarity distance between objects i and j of $[n]$ by

$$D_{ij}^* = \left(\frac{1}{n} \sum_{\ell=1}^n [F_{\pi_i \ell} - F_{\pi_j \ell}]^2 \right)^{1/2}, \quad (11)$$

which is the counterpart of the neighborhood distance for graphon analysis – see [Lovász, 2012]. Intuitively, D_{ij}^* is expected to be small when π_i and π_j are close, while D_{ij}^* is expected to be large when π_i and π_j are distant.

If we knew in advance the variances of the E_{ij} 's, we could simply estimate $(D^*)_{ij}^2$ by the unbiased estimator $n^{-1} [\|A_i - A_j\|^2 - (\sum_{k=1}^n \text{var}(E_{ik}) + \text{var}(E_{jk}))]$. However, in most situations (e.g. binary data), the variances of noise terms are unknown and it is not possible to craft an unbiased estimator of D_{ij}^* or $(D^*)_{ij}^2$. The problem of estimating D_{ij}^* has been studied in [Issartel, 2021], on which we rely to produce the estimator \widehat{D} (Algorithm 4).

The precise definition of the procedure `DistanceEstimation` to compute \widehat{D} is postponed to appendix B. We only give the general idea here. The quadratic form $n(D_{ij}^*)^2$ can be decomposed as $\sum_{\ell=1}^n F_{\pi_i \ell}^2 + \sum_{\ell=1}^n F_{\pi_j \ell}^2 - 2 \sum_{\ell=1}^n F_{\pi_j \ell} F_{\pi_i \ell}$. An unbiased estimator of the cross term is simply $\sum_{\ell=1}^n A_{j\ell} A_{i\ell}$. However, the quadratic term $\sum_{\ell=1}^n F_{\pi_i \ell}^2$ is more challenging to handle. In `DistanceEstimation`, we estimate this quantity by a cross-term $\sum_{\ell=1}^n A_{i\ell} A_{\widehat{m}_i \ell}$, where the data-driven object \widehat{m}_i is chosen in such a way that $D_{i\widehat{m}_i}^*$ is as small as possible. Since $D_{i\widehat{m}_i}^*$ is unknown (and still to be estimated), we rely on a specific procedure to choose \widehat{m}_i –see [Issartel, 2021] or appendix B.1.

Our interest in the similarity distance matrix comes from the fact that this distance D_{ij}^* is somewhat related to the distance $|\pi_i - \pi_j|$. More precisely, we hope that D_{ij}^* is small when π_i is close to π_j and that D_{ij}^* is large when π_i is far from π_j . Unfortunately, this is not true for general Robinson matrices F . However, if we make additional assumptions on F , such as Assumption 2.1, we are able to establish some formal connection between the two distances –see Lemma C.1 in appendix.

3.2 Rough estimation of $H^{(\pi)}$ by aggregation of local bisections

In the second step `AgregLocalBisection` of SALB, we build a rough estimator of $H^{(\pi)}$ from the estimator \widehat{D} of the distance matrix. The procedure summarized in Algorithm 2 works as follows. For each object $i \in [n]$, we first build two subsets $G_i^{(1)}$ and $G_i^{(2)}$ of $[n]$ such that, with respect to the oracle ordering π , all objects in $G_i^{(1)}$ are on one side of i and all objects of $G_i^{(2)}$ are on the other side of i . This is the purpose of the corresponding function `LocalBisection` – see Line 2 of Algorithm 2 – which will be described after Algorithm 2. Then, we align all sets $(G_i^{(1)}, G_i^{(2)})$ in a coherent manner by deciding which ones are on the left of i , and which ones are on the right of i . The corresponding function `Orientation` – Line 4 of Algorithm 2 – is postponed to appendix B.1 (pseudo code in Algorithm 6). This common orientation allows us to obtain the collection of subsets $(L, R) := (L_i, R_i)_{i \in [n]}$ of nodes that are left (resp. right) of i . Finally, in Lines 5–12, we simply build upon the (L_i, R_i) 's to define a comparison matrix H . We set $H_{ij} = -1$ if $i \in L_j$ or $j \in R_i$, and set $H_{ij} = 1$ if $i \in R_j$ or $j \in L_i$. If none of these conditions is met, we simply keep $H_{ij} = 0$.

Algorithm 2 AgregLocalBisection

Require: $(D, \delta_1, \delta_2, \delta_3)$ **Ensure:** $H \in \{-1, 0, 1\}^{n \times n}$

```
1: for  $i = 1, \dots, n$  do
2:    $(G_i^{(1)}, G_i^{(2)}) = \text{LocalBisection}(i, D, \delta_1, \delta_2, \delta_3)$ 
3: end for
4:  $(L, R) = \text{Orientation}(G_1^{(1)}, G_1^{(2)}, \dots, G_n^{(1)}, G_n^{(2)})$ 
5: Set  $H = 0_{n \times n}$ 
6: for  $i, j = 1, \dots, n$  do
7:   if  $i \in L_j$  or  $j \in R_i$  then
8:      $H_{ij} = -1$ 
9:   else if  $i \in R_j$  or  $j \in L_i$  then
10:     $H_{ij} = 1$ 
11:   end if
12: end for
```

Let us describe how `LocalBisection` (Line 2) builds the subset $G_i^{(1)}$ and $G_i^{(2)}$ from a distance matrix estimate D . For $i \in [n]$ and $\delta = (\delta_1, \delta_2, \delta_3)$, define the graph $\mathcal{G}_{i,\delta}$ of node set $[n] \setminus \{i\}$ as follows:

$$\text{Put an edge between any nodes } k \text{ and } l \text{ if, } D_{kl} \leq \delta_1 \text{ and } D_{ik} \vee D_{il} \geq \delta_2. \quad (12)$$

Then, the result $(G_i^{(1)}, G_i^{(2)}) = \text{LocalBisection}(i, D, \delta_1, \delta_2, \delta_3)$ is any two largest connected components of $\mathcal{G}_{i,\delta}$ that contain at least a node k such that $D_{ik} \geq \delta_3$. By convention, $G_i^{(1)}$ or $G_i^{(2)}$ can be empty sets if there exist less than two such connected components. The pseudo code of `LocalBisection` is written in Algorithm 5.

Let us explain the rationale behind the construction of $(G_i^{(1)}, G_i^{(2)})$. Recall that, for a bi-Lipschitz matrix F , the distance D_{ij}^* behaves analogously to $|\pi_i - \pi_j|$ (see Lemma C.1 in appendix) and assume, for the purpose of this discussion, that we have perfectly estimated D^* so that $D = D^*$ above. Then, for suitable tuning parameters δ_1 and δ_2 , any node (k, l) such that $|\pi_k - \pi_l| = 1$ and $|\pi_k - \pi_i|$ is large will be connected in $\mathcal{G}_{i,\delta}$ by (12) since this implies that D_{kl} is small and D_{ik} is large. Besides, any two nodes (k, l) which are not on the same side of i with respect to π cannot be connected. (Indeed, either $|\pi_k - \pi_l|$ is large but this enforces that D_{kl} is large which prohibits k and l to be connected in the construction (12), or $|\pi_k - \pi_l|$ is small but this enforces that $D_{ik} \vee D_{il}$ is small which prevents again k and l to be connected in (12).) In conclusion, we expect the graph $\mathcal{G}_{i,\delta}$ to have two connected components which contain, respectively, all nodes that are far from i on the left of i (w.r.t π), and all nodes that are far from i on the right of i .

Discussion of the tuning parameters $\delta = (\delta_1, \delta_2, \delta_3)$. For the purpose of this discussion, let us assume that F is bi-Lipschitz as in the previous section, and that the quantities α , β , and σ are considered as fixed positive constants. In this case, we state in Lemma C.1 that $nD_{kl}^*/|\pi_k - \pi_l| \in (\alpha/2, \beta)$ as long as k and l are close enough, namely $|\pi_k - \pi_l| < n/4$ or $D_{kl}^* < 1/4$. Besides, Lemma D.1 in appendix implies that, with high probability, $|D_{kl}^* - \widehat{D}_{kl}|$ is at most of the order $(\sqrt{\log(n)}/n)^{1/4}$ uniformly for all k, l close enough. Thus, when they are small enough, the distances \widehat{D}_{kl} and D_{kl}^* and $|\pi_k - \pi_l|/n$ are equivalent, up to constant factors

α, β and to additive error of the order $(\sqrt{\log(n)}/n)^{1/4}$. Going back to the construction (12) of the graph $\mathcal{G}_{i,\delta}$, we first need to choose the tuning parameter δ_1 , in such a way that, consecutive points in the ordering π are connected in $\mathcal{G}_{i,\delta}$, that is for k and l satisfying $|\pi_k - \pi_l| = 1$, we have $\widehat{D}_{kl} \leq \delta_1$ with high probability. As a consequence, taking $\delta_1 = n^{-1/5}$, that is slightly larger than $(\sqrt{\log(n)}/n)^{1/4}$, is sufficient for our purpose. Then, we choose δ_2 in order to preclude the existence of any edge between any two nodes k and l that lie on both sides of i . If we set $\delta_2 = \log(n)\delta_1 = \log(n)n^{-1/5}$, then $\widehat{D}_{ij} \vee \widehat{D}_{ik} \geq \delta_2$ implies that $|\pi_k - \pi_i| \vee |\pi_l - \pi_i|$ is at least of the order of $n\delta_2$. As a consequence, if k and l are on two different sides of i (w.r.t. π), this implies that $|\pi_k - \pi_l|$ is at least of the order $n\delta_2$ and we cannot have \widehat{D}_{kl} smaller than δ_1 , hence an absence of edge between k and l in $\mathcal{G}_{i,\delta}$ by construction (12). For this choice of δ_2 , each connected component of $\mathcal{G}_{i,\delta}$ is thus made of nodes located on only one side of i (w.r.t. π). Finally, we can fix $\delta_3 = \log^2(n)\delta_1 \gg \delta_2$, so that the two chosen connected components $(G_i^{(1)}, G_i^{(2)}) = \text{LocalBisection}(i, D, \delta_1, \delta_2, \delta_3)$ contain all the points l such that $|\pi_l - \pi_i|$ is at least of the order of δ_3 (up to multiplicative constants depending on α, β, σ). These guarantees are formalized in appendix D.3.1.

3.3 Local refinement of the estimation of $H^{(\pi)}$

In this step, we build a second matrix \widehat{H}_2 whose support does not intersect that of \widehat{H}_1 . The matrix \widehat{H}_1 built in the previous step possibly contains a lot of zero entries. The purpose of this local refinement is to compare all those objects i and j such that $(\widehat{H}_1)_{ij} = 0$. For each such (i, j) , the broad idea is to test the relative position of i and j by relying on the observation A and our present knowledge \widehat{H}_1 of the ordering. As there are complex dependencies between \widehat{H}_1 and A , the actual procedure `LocalRefineWS` involves sample splitting scheme. For the sake of clarity, we describe the simpler procedure `LocalRefine` in Algorithm 3 without any sample splitting and postpone `LocalRefineWS` to the appendix H.

Let us explain how we compare the relative position of i and j . Define L_{ij}^* (resp. $R_{i,j}^*$) as the set of objects k such that $\pi_k \leq \pi_i \wedge \pi_j$ (resp. $\pi_k \geq \pi_i \vee \pi_j$). If $\pi_i < \pi_j$, then we deduce from the Robinson property (1) of the matrix F that $\sum_{k \in L_{ij}^*} F_{\pi_i \pi_k} - F_{\pi_j \pi_k}$ is positive when $L_{ij}^* \neq \emptyset$. Similarly, we also have that $\sum_{k \in R_{i,j}^*} F_{\pi_j \pi_k} - F_{\pi_i \pi_k}$ is positive. Obviously, we do not know L_{ij}^* and $R_{i,j}^*$, but we can estimate them thanks to \widehat{H}_1 . More specifically, given any comparison matrix H_1 , we define $L_{ij} = \{k, (H_1)_{ki} = -1 \text{ and } (H_1)_{kj} = -1\}$ and $R_{i,j} = \{k, (H_1)_{ki} = 1 \text{ and } (H_1)_{kj} = 1\}$. On the event where H_1 does not make any false comparison, L_{ij} is a subset¹ of L_{ij}^* and $R_{i,j}$ is a subset of $R_{i,j}^*$. Then, we consider the statistics $l_{ij} = \sum_{k \in L_{ij}} A_{ik} - A_{jk}$ and $r_{ij} = \sum_{k \in R_{i,j}} A_{ik} - A_{jk}$. If $|l_{ij}| \vee |r_{ij}| < 5\sigma\sqrt{n \log(n)}$, then we cannot draw a statistically significant conclusion. Otherwise, when either $|l_{ij}|$ or $|r_{ij}|$ is higher than $5\sigma\sqrt{n \log(n)}$, we set H_{ij} to -1 or 1 depending on the sign l_{ij} or r_{ij} – see Lines 8 and 11 in Algorithm 3.

¹Actually, as π is not identifiable, we can also have L_{ij} is a subset of $R_{i,j}^*$ and $R_{i,j}$ is a subset of L_{ij}^* but this is not an issue for our purpose.

Algorithm 3 LocalRefine

Require: (H_1, A, σ) **Ensure:** $H \in \{-1, 0, 1\}^{n \times n}$

```
1: Initiate  $H = 0_{n \times n}$ 
2: for  $(1 \leq i < j \leq n)$  s.t.  $(H_1)_{ij} = 0$  do
3:    $L_{ij} = \{k, (H_1)_{ki} = -1 \text{ and } (H_1)_{kj} = -1\}$ 
4:    $R_{ij} = \{k, (H_1)_{ki} = 1 \text{ and } (H_1)_{kj} = 1\}$ 
5:    $l_{ij} = \sum_{k \in L_{ij}} A_{ik} - A_{jk}$ 
6:    $r_{ij} = \sum_{k \in R_{ij}} A_{ik} - A_{jk}$ 
7:   if  $|l_{ij}| \geq 5\sigma\sqrt{n \log(n)}$  then
8:      $H_{ij} = -\text{sign}(l_{ij})$  and  $H_{ji} = -H_{ij}$ 
9:   else
10:    if  $|r_{ij}| \geq 5\sigma\sqrt{n \log(n)}$  then
11:       $H_{ij} = \text{sign}(r_{ij})$  and  $H_{ji} = -H_{ij}$ 
12:    end if
13:  end if
14: end for
```

4 Beyond bi-Lipschitz matrices via average type constraints

Bi-Lipschitz matrices (Assumption 2.1) offer a simple model that facilitates exposition, but that may be quite restrictive in applications. We therefore present more realistic assumptions in this section, which are at the core of our analysis of SALB. They seem (to us) somewhat natural and minimal assumptions for SALB's performance, and their presentation in section 4.1 makes our results more transparent. We also illustrate in section 4.2 the flexibility of these new assumptions compared to bi-Lipschitz matrices.

4.1 Seriation error under average type assumptions on F

To prove good performances for the distance based method – `AgregLocalBisection` (Algorithm 2) – we need the input $D = [D_{ij}]$ to be informative on the distances $|\pi_i - \pi_j|$ in some way. Specifically, we will need D_{ij} to be locally equivalent to $|\pi_i - \pi_j|/n$ as in (13).

Assumption 4.1 (Local Distance Equivalence). *For any constants $0 < \tilde{\alpha} \leq \tilde{\beta}$, $0 < r$ and $0 \leq \omega$, let $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ be the collection of symmetric matrices $D \in \mathbb{R}^{n \times n}$ that satisfy*

$$\tilde{\alpha} \frac{|\pi_i - \pi_j|}{n} - \omega \leq D_{ij} \leq \tilde{\beta} \frac{|\pi_i - \pi_j|}{n} + \omega \quad (13)$$

for all $i, j \in [n]$ such that $\frac{|\pi_i - \pi_j|}{n} \wedge D_{ij} \leq r$.

Thus, a matrix D satisfies Assumption 4.1 if its entries D_{ij} are lower and upper bounded by the distances $|\pi_i - \pi_j|/n$, up to factors $\tilde{\alpha}$ and $\tilde{\beta}$, and to an additive error ω . This condition is local as it only concerns the pairs i, j within a distance r . Assumption 4.1 is therefore quite general, since it enforces no constraints for medium and large distances, and it permits bounded distortions of small distances (by $\tilde{\alpha}$ and $\tilde{\beta}$), and even gives some slack (by $\pm\omega$). In contrast with usual properties of distances, we assume no form of transitivity, e.g. $\pi_i - \pi_k \geq \pi_j - \pi_k \implies D_{ik} \geq D_{jk}$, and no form of triangular inequality.

While Assumption 4.1 is at the core of our analysis of `AgregLocalBisection` for the input $D = \widehat{D}$, we need another type of assumption for proving good performances for `LocalRefine` and `LocalRefineWS`. To analyze the statistical test in Algorithm 3, we will use Assumption 4.2 which ensures that, for any (close) indices i and j , there is a separation between the extreme similarities of i and that of j (i.e. between the sums of their pairwise similarities $F_{i\ell}$ and $F_{j\ell}$ for ℓ 's running over the left or right sets of i, j).

Assumption 4.2 (Separated Cumulative Similarities). *For any constants $0 < \gamma, r', r''$, let $\mathcal{SCA}[\gamma, r', r'']$ be the collection of matrices that satisfy, for all $i < j$ and $|i - j| \leq r'n$,*

$$\begin{aligned} \sum_{\ell: 1 \leq \ell < i - r''n} F_{i\ell} - F_{j\ell} &\geq \gamma|i - j| && \text{if } \frac{i}{n} \geq \frac{1 - r'}{2}, \\ \sum_{\ell: j + r''n < \ell \leq n} F_{j\ell} - F_{i\ell} &\geq \gamma|i - j| && \text{if } \frac{j}{n} \leq \frac{1 + r'}{2}. \end{aligned} \quad (14)$$

The constant γ represents the cumulative signal when comparing the sum of similarities of $i < j$ over their left set $\{\ell : \ell < i\}$ or right set $\{\ell : j < \ell\}$. (For technical reasons, however, Assumption 4.2 involves instead the left and right subsets reduced by a length r'' .) Assumption 4.2 is local as it only concerns the pairs i, j at distance less than r' . This is sufficient for our purpose, since `LocalRefineWS` only deals with the pairs i, j left undetermined by `AgregLocalBisection` and, as will be proved later, such i, j are within a small distance from each other (with high probability).

Let us give a justification for Assumption 4.2. To be able to compare i, j , there should be at least one set in $[n]$ on which the similarities F_{iS} and F_{jS} of i, j are different; otherwise, the estimation of the comparison $H_{ij}^{(\pi)}$ is hopeless. Then, since F has a Robinson structure (1), a natural choice of discriminating set for $i < j$ is either the left set $\{\ell : \ell < i\}$, or the right set $\{\ell : j < \ell\}$ of i and j .

Theorem 4.1 ensures that our estimator $\hat{\pi}_o$ still achieves the same $\sqrt{\log(n)/n}$ rate, when the distance matrix D^* in (11) belongs to $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ and when $F \in \mathcal{SCA}[\gamma, r', r'']$. We simplify the statement of the next result by setting the tuning parameters $(\delta_1, \delta_2, \delta_3)$ of Algorithm 1 to $\delta_1 = n^{-1/5}$, $\delta_2 = \log(n)\delta_1$, and $\delta_3 = \log(n)\delta_2$. Theorem 4.1 is proved in Appendix C.6.

Theorem 4.1. *For $\tilde{\alpha}, \tilde{\beta}, r, \sigma, \gamma, r', r'' > 0$, there exists a positive constant $C_{\tilde{\alpha}, \tilde{\beta}, r, \sigma, r', r''}$ only depending on $(\tilde{\alpha}, \tilde{\beta}, r, \sigma, r', r'')$ and a numerical constant $C > 0$ such that the following holds for any $n \geq C_{\tilde{\alpha}, \tilde{\beta}, r, \sigma, r', r''}$ and for any matrix $F \in \mathcal{SCA}[\gamma, r', r'']$ such that $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$. With probability higher than $1 - 9/n^2$, the permutation $\hat{\pi}_o$ computed by SALB satisfies*

$$L_{\max}(\hat{\pi}_o, \pi) \leq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}. \quad (15)$$

In other words, the polynomial-time algorithm SALB estimates the position of each object up to an error of the order $\frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}$. In particular, one sees that the error becomes smaller for larger γ . Theorem 4.1 is stated for simplicity in a setting where α, β , and σ are constants and n is relatively large, but SALB achieves the convergence rate (15) even in settings where the quantities α, β , and σ depend on n . We refer to Theorem C.4 for a general bound.

In comparison with Theorem 2.1, the bound (15) involves the constant γ of Assumption 4.2 instead of the constant α of Assumption 2.1. In fact, Theorem 2.1 is particular case of Theorem 4.1, since the assumptions of the latter hold for any bi-Lipschitz F . More precisely, if $F \in \mathcal{BL}[\alpha, \beta]$, then $F \in \mathcal{SCA}[\gamma, r', r'']$ for $\gamma = \alpha/4$, and $D^* \in \mathcal{LDE}[\alpha/2, \beta, 0, r]$; see Lemma C.1 and C.2 for details.

Remark 4.1. Although Assumptions 4.1 and 4.2 are distinct hypotheses, they are not independent. As said above, if $F \in \mathcal{BL}[\alpha, \beta]$, then Assumption 4.1 and 4.2 hold for $\tilde{\alpha} = \alpha/2$ and $\gamma = \alpha/4$, so we have the strong relation $\tilde{\alpha} = \gamma/2$ between Assumption 4.1 and 4.2.

4.2 Motivation for average type Assumption 4.1 and 4.2

The merits of the bi-Lipschitz Assumption 2.1 was to offer a simple model, but this assumption is sometimes too restrictive. In particular, it enforces “long range similarities” which are unrealistic in applications where distant objects have a small similarity (see next paragraph). Besides being more transparent of our analysis, Assumption 4.1 and 4.2 alleviate this long range issue; see the last paragraph for an example. This relaxation comes from the following: the bi-Lipschitz assumption is an entry-wise constraint on the matrix F , while Assumption 4.1 and 4.2 are only average constraints on F 's columns (respectively an l_2 and l_1 type average constraints). Thus, Assumption 4.1 and 4.2 offer more flexibility to fit data in applications.

Long-range affinity: Taking objects i, j, k that are far from each other, say $i = 1$, $j = n/2$ and $k = n$, Assumption 2.1 enforces the variation $n|F_{1n} - F_{\frac{n}{2}n}| \asymp_{\alpha, \beta} n$. Such a long-range constraint is not satisfied in applications where the pairwise similarity between distant objects is nearly zero (e.g. $F_{1n} = F_{\frac{n}{2}n} = 0$), and only close objects have significant similarities.

Example of a Robinson matrix F that is not bi-Lipschitz, but satisfies Assumption 4.1 and 4.2: Given some $k \in [n/2]$, let F such that $F_{ij} = a_{|i-j|}$ where $a_i = [(k-i) \vee 0]/n$ for $i = 0, \dots, n-1$. Then, F is a monotone Toeplitz matrix, which is a standard model of Robinson matrix (Example 1 in appendix A). Obviously, F violates the long-range affinity of the bi-Lipschitz Assumption 2.1. But, one can readily check that Assumption 4.1 and 4.2 are satisfied, that is there exist positive constants $\tilde{\alpha}, \tilde{\beta}, r$ and γ, r, r'' such that the distance matrix D^* in (11) belongs to the set $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$, and F is in $\mathcal{SCA}[\gamma, r', r'']$. Besides, when k is proportional to n , say $k = n/2$, one can see that $\tilde{\alpha}, \tilde{\beta}$ and γ are numerical constants.

5 Beyond bi-Lipschitz matrices via latent space formulation

We relax the bi-Lipschitz assumption in two different ways: (i) by removing the entry-wise constraints and replacing them with average constraints on the matrix F ; we did this in section 4. (ii) by removing the homogeneity assumption that all F 's columns have a similar shape, in particular that consecutive columns F_i and F_{i+1} are almost identical. We are about to explore this second way, using a latent space generalization of the former bi-Lipschitz assumption. Our result is stated in section 5.1, and the motivation for this generalization is discussed in section 5.2.

5.1 Seriation error on latent bi-Lipschitz matrices

We replace the former bi-Lipschitz assumption by Assumption 5.1, which consists in replacing the regular grid $\{i/n\}_{i \in [n]}$ by general (unknown) points $\{y_i\}_{i \in [n]}$ of $[0, 1]$.

Assumption 5.1 (Latent bi-Lipschitz Matrix). *For any points $y_1 < \dots < y_n$ in $[0, 1]$, and constants $0 < \alpha \leq \beta$, let $\mathcal{L}[(y_i), \alpha, \beta]$ be the collection of matrices $F \in \mathbb{R}^{n \times n}$ that satisfy*

$$|F_{ik} - F_{jk}| \leq \beta |y_i - y_j| \quad \text{for all } (i, j, k) ; \quad (16)$$

$$F_{ik} - F_{jk} \geq \alpha |y_i - y_j| \quad \text{for all } k < i < j ; \quad (17)$$

$$F_{jk} - F_{ik} \geq \alpha |y_i - y_j| \quad \text{for all } i < j < k .$$

If the unknown points y_i are completely arbitrary in $[0, 1]$, the seriation problem is not well defined – see Remark 5.1.2. For this purpose, we consider the additional conditions (18-19), which ensure that the y_1, \dots, y_n are nearly well spread in $[0, 1]$. Compared to the regular grid ($y_i = i/n$), the conditions (18-19) still give significant liberty to the y_i 's – see section 5.2 for a discussion.

Unclustered spreading: Let $\eta > 0$ such that

$$\sup_{y \in [0,1]} \min_{i \in [n]} |y - y_i| \leq \eta . \quad (18)$$

This means that, for any interval $I \subset [0, 1]$ of length η , there is (at least) one $y_i \in I$.

Balanced spreading: Let $\tilde{\eta} \in (0, 1)$ such that

$$\text{Card}\left\{i \in [n] : y_i \in \left[0, \frac{1}{4}\right]\right\} \geq \tilde{\eta} n , \quad \text{Card}\left\{i \in [n] : y_i \in \left[\frac{3}{4}, 1\right]\right\} \geq \tilde{\eta} n . \quad (19)$$

In words, the extreme intervals of length $1/4$ contains (at least) a proportion $\tilde{\eta}$ of the y_i 's.

Remark 5.1. on modeling assumptions

1. The constant η in (18) is non-increasing with n . In standard latent space models, the y_i 's are often assumed to come from a uniform sample of $[0, 1]$, and hence η goes to zero as $n \rightarrow \infty$. By contrast, we make no assumption on the decay of η (with n) in this paper; the y_i 's are deterministic. (If the y_i 's were random, our results should be stated conditionally to the randomness of the y_i 's.) The proximity condition (18), which bounds the distance between consecutive objects y_i, y_{i+1} , is crucial for our distance based method – `AgregLocalBisection` – which operates locally (on small distances).
2. Without (18-19), the y_i 's could be clustered around a small number of points (that are properly spaced) in $[0, 1]$, hence the problem would become a clustering problem rather than a seriation problem, and our algorithm is not suited for clustering data. The question of handling simultaneously clustering and seriation is interesting, but is beyond the scope of the paper.

To work in a latent setting as in Assumption 5.1, it is crucial to introduce a more general loss than the L_{\max} loss, because L_{\max} is too crude for assessing estimator performances in latent space models – see Remark 5.2.1. Given $x_i = y_{\pi_i}$ for all i , with $\pi = (\pi_i)_{i \leq n}$, we define a new loss, denoted L_{comp} , as follows. For $\epsilon > 0$, it is bounded by $L_{\text{comp}}(\hat{\pi}, \pi) \leq \epsilon$ if there exists $s \in \{\pm\}$ such that

$$\forall i, j \text{ s.t. } |x_i - x_j| \geq \epsilon : \quad H_{ij}^{(\hat{\pi})} = s H_{ij}^{(\pi)} , \quad (20)$$

where we used the definition (10) for the map $\tilde{\pi} \mapsto H_{ij}^{(\tilde{\pi})}$ which sends any permutation $\tilde{\pi}$ to a comparison matrix $H_{ij}^{(\tilde{\pi})}$. The L_{comp} was studied in [Janssen and Smith, 2020]. It returns the minimal distance $|x_i - x_j|$ above which, all comparisons $\hat{\pi}_i < \hat{\pi}_j$ coincide with the true comparisons $\pi_i < \pi_j$. The sign change $s \in \{\pm\}$ comes from the identifiability in the seriation problem which holds only up to a reversal of the ordering π . The L_{comp} loss is a natural extension of L_{max} and is even equivalent to L_{max} under the original Assumption 2.1; see Remark 5.2.2.

Remark 5.2. on the L_{comp} -loss:

1. Since L_{max} returns the maximum of all estimation errors of the π_i , $i \in [n]$, it only reflects the learning difficulty of the worst π_i among the π_1, \dots, π_n . For this crude measure, no estimator is consistent for the L_{max} loss in the latent setting (Assumption 5.1). As an extreme example, let $y_1 = \dots = y_k$ for $k \in [n]$, then their similarity vectors are all equal $F_{\pi_1} = \dots = F_{\pi_k}$ (by Assumption 5.1). There is therefore no hope of recovering the positions π_1, \dots, π_k . Inevitably, the L_{max} loss of any estimator is at least of the order of k/n . Thus, L_{max} only reflects the learning impossibility of these k (identical) positions.
2. In the original matrix Assumption 2.1, the L_{comp} and L_{max} losses are equivalent (up to a factor 2). We have $L_{\text{max}}(\hat{\pi}, \pi) \leq 2\epsilon$ when $L_{\text{comp}}(\hat{\pi}, \pi) \leq \epsilon$ (by taking $H = H^{\hat{\pi}}$ in Proposition C.9 and using $\hat{\pi} = \pi^{H^{\hat{\pi}}}$). Vice versa, we can readily check that $L_{\text{comp}}(\hat{\pi}, \pi) \leq 2\epsilon$ when $L_{\text{max}}(\hat{\pi}, \pi) \leq \epsilon$. This equivalence between L_{comp} and L_{max} in bi-Lipschitz matrices (Assumption 2.1) is not true anymore in latent bi-Lipschitz matrices (Assumption 5.1).

The next theorem shows that our estimator $\hat{\pi}_o$ performs well even in latent bi-Lipschitz matrices, when the latent y_i 's satisfy the conditions (18-19). To simplify the statement of the next result, we consider the special case of a clustering constant η in (18) that converges to zero, that is $\eta := \eta_n \rightarrow 0$ as $n \rightarrow \infty$, and we set the tuning parameters $(\delta_1, \delta_2, \delta_3)$ in Algorithm 1 to $\delta_1 = n^{-1/5} + \eta_n^{1/3}$, $\delta_2 = -\log(\delta_1)\delta_1$, and $\delta_3 = -\log(\delta_1)\delta_2$,

Theorem 5.1. *For $\alpha > 0$, $\beta > 0$, $\sigma > 0$ and any sequence $\bar{\eta} = (\eta_n)_{n \geq 1}$ such that $\eta := \eta_n \rightarrow 0$, there exists a positive constant $C_{\alpha, \beta, \sigma, \bar{\eta}}$ only depending on $(\alpha, \beta, \sigma, \bar{\eta})$ and a numerical constant $C > 0$ such that the following holds for any $n \geq C_{\alpha, \beta, \sigma, \bar{\eta}}$, for any $\tilde{\eta} > 0$ and $y_1, \dots, y_n \in [0, 1]$ complying with (18-19), and for any $F \in \mathcal{LB}\mathcal{L}[(y_i), \alpha, \beta]$. With probability higher than $1 - 9/n^2$, the permutation $\hat{\pi}_o$ computed by SALB satisfies*

$$L_{\text{comp}}(\hat{\pi}_o, \pi) \leq C \frac{\sigma}{\tilde{\eta}\alpha} \sqrt{\frac{\log(n)}{n}} . \quad (21)$$

In comparison with Theorem 2.1, the rate (21) also contains the term $\tilde{\eta}$ of condition (19), which measures the balance of y_i 's spreading in $[0, 1]$. This highlights the effect of well spread points y_1, \dots, y_n in $[0, 1]$ for the latent bi-Lipschitz setting (Assumption 5.1). Note that the rate in Theorem 2.1 follows directly from (21) since $\tilde{\eta} = 1/4$ when $y_i = i/n$.

Theorem 5.1 is stated in a special case setting where $\eta \rightarrow 0$ as $n \rightarrow \infty$, (and the α, β, σ are considered as constants whereas n is large). But SALB achieves the convergence rate (21) even in settings where η does not converge to zero (and the quantities α, β, σ may depend on n), but the choices of the tuning parameters $\delta_1, \delta_2, \delta_3$ are more intricate. We refer to appendix L.1 for a general bound.

Assumption 5.1 is similar to the latent space model considered in [Giraud et al., 2021]. The (non-efficient) procedure in [Giraud et al., 2021] attains the rate $\sqrt{\log(n)/n}$ in the special case of the regular grid $y_i = i/n$. By contrast, our estimator $\hat{\pi}_o$ achieves the rate $\sqrt{\log(n)/n}$ even when the latent points (y_i) depart from the regular grid (i/n) , and even when two consecutive points y_i, y_{i+1} are at a constant distance from each other (see the general theorem in appendix L.1). This improvement is especially interesting for applications where one wants to order objects that are non evenly spread in some feature space. Our improvements over the earlier work [Giraud et al., 2021] also include a rate (21) giving explicit dependencies in the problem parameters, and an optimal estimator $\hat{\pi}_o$ over the class of latent bi-Lipschitz matrices that is efficient (i.e. whose time complexity is polynomial in n). The proof of Theorem 5.1 is in appendix L.

[Janssen and Smith, 2020, Giraud et al., 2021, Natik and Smith, 2021] prove theoretical guarantees for efficient procedures in Toeplitz matrices (Example 1 in appendix A), which are generated by geometric latent space models (Example 3 in appendix A). Closer to our paper, in latent bi-Lipschitz matrices [Giraud et al., 2021] show that, if the matrix is approximately a Toeplitz matrix, then the standard spectral procedure coupled with a post-processing step achieves the (optimal) error bound $L_{\text{comp}}(\hat{\pi}, \pi) \leq C\sqrt{\log(n)/n}$ with high probability. Although their rate is of the same order, their assumptions are more restrictive than ours. Further from our paper, [Janssen and Smith, 2020, Natik and Smith, 2021] study the special case of network data $A \in \{0, 1\}^{n \times n}$ generated by the popular graphon model, where the latent points (y_i) are a uniform sample / the regular grid of $[0, 1]$ (it is a special case of Example 2 in appendix A). On the one hand, [Janssen and Smith, 2020] assumes a Toeplitz matrix (“uniformly embedded graphon” with a uniform sample), a constant signal $F_{ij} = c$ beyond a certain distance $|i - j|$, as well as a technical condition on the square of the graphon, to guarantee that the thresholded squared adjacency matrix $A^{(2)}$ used in their algorithm is close to a Robinson matrix. Their new procedure, which is based on a thresholded version of $A^{(2)}$, provably achieves the error bound $L_{\text{comp}}(\hat{\pi}, \pi) \leq \frac{(\log(n))^5}{\sqrt{n}}$ with high probability. On the other hand, [Natik and Smith, 2021] assumes a Toeplitz matrix (“uniformly embedded graphon” with a uniform sample) and some \mathcal{C}^1 -smooth graphon with a strictly negative derivative (called “nice” graphon); they show that the standard spectral algorithm (with post-processing) attains a similar error bound than above. Although the rates in [Janssen and Smith, 2020, Natik and Smith, 2021] are similar (up to log factors), their assumptions are not directly comparable to ours. We mention that [Janssen and Smith, 2020, Natik and Smith, 2021] also propose relaxations of the aforementioned assumptions, but these are rather technical and difficult to interpret here.

5.2 Motivation for the latent space model

While the simple bi-Lipschitz Assumption 2.1 facilitated our exposition, it is quite restrictive in practice, in particular because of the homogeneity assumption on F ’s columns, which fails to fit heterogeneous data in applications. By comparison, the latent space formulation (Assumption 5.1), which still offers a simple model to interpret, is flexible enough to alleviate the homogeneity issue, by allowing more variations between F ’s columns.

Interpretation. Setting $x_i = y_{\pi_i}$, Assumption 5.1 has the following simple interpretation. Each object i has an unknown feature x_i and a similarity vector $\mathbb{E} A_i = F_{\pi_i}$. Two objects i, j with close features x_i, x_j in $[0, 1]$ will have almost identical similarity vectors, while two

objects with distant features will have very different similarity vectors. In this setting, the task of finding π is equivalent to reordering the x_i 's in the latent space $[0, 1]$.

Breaking the homogeneity restriction? In bi-Lipschitz matrices (Assumption 2.1), two consecutive columns are almost identical since $|F_{ik} - F_{(i+1)k}| \asymp_{\alpha,\beta} n^{-1}$. Therefore, the (squared) Euclidean distance between consecutive vectors equals almost zero: $\|F_i - F_{i+1}\|_2^2 \asymp_{\alpha,\beta} n^{-1}$. (Here, we used $F_{i(i+1)} = F_{(i+1)i}$, and considered the vectors F_i and F_{i+1} in \mathbb{R}^{n-1} , by removing their respective null coordinates F_{ii} and $F_{(i+1)(i+1)}$.) This homogeneity restriction between consecutive columns is relaxed in latent bi-Lipschitz matrices (Assumption 5.1) where one may have $y_{i+1} - y_i = \eta$ for a positive constant $\eta > 0$, and hence the (squared) Euclidean distance $\|F_i - F_{i+1}\|_2^2 \asymp_{\alpha,\beta} \eta^2 n$ may diverge. Compared to the null distance in bi-Lipschitz matrices, latent bi-Lipschitz matrices thus offer a significant relaxation of the homogeneity constraint. Hopefully, such a latent space model would fit better heterogeneous data encountered in applications, e.g. in networks where popular individuals have many interactions (i.e. many high similarities), while some others have much fewer interactions (many low similarities).

6 Discussion

We studied the seriation problem under bi-Lipschitz assumptions on the signal matrix F , and focused on the L_{\max} loss which returns the maximum estimation error of all positions π_1, \dots, π_n . The good news is that, even for the crude loss L_{\max} , in such a general and unspecified model as bi-Lipschitz matrices, we successfully characterized the optimal rate of seriation, as a function of the problem parameters. We also gave a polynomial time estimator that achieves this optimal rate. Besides, we proved that the estimator also enjoys good performances in more general sets of matrices, namely, matrices with only average type constraints, and latent bi-Lipschitz matrices. Overall, our work showcases the versatility of this permutation estimator.

As a preliminary step, we estimated the measure D^* defined in (11), via the estimator \widehat{D} described in appendix B.1. In principle, we could have used any other (good) estimator of D^* , or even any other measures than D^* , as long as this measure is informative on the latent distances $|\pi_i - \pi_j|/n$. Perhaps surprisingly, this measure does not have to fulfill the usual properties of distances and linear orderings, such as the triangular inequality or the transitivity. Thus, our approach is quite general, and perhaps could be exported elsewhere.

One might hope that the distances D_{ij}^* and $|\pi_i - \pi_j|/n$ are almost the same, or sufficiently similar for trying to recover π from D^* directly. This will not work in general, because both distances may behave very differently. For example, D_{ij}^* can be a huge distortion of $|\pi_i - \pi_j|/n$, and this distortion can go both ways (contraction or dilation). Thus, the two distances are sometimes in contradiction in the following sense. One can find R-matrices F such that $D_{1n}^* < D_{1(n/2)}^*$ while the reverse holds for the ordering, i.e., $|\pi_1 - \pi_n| > |\pi_1 - \pi_{n/2}|$. More generally, D^* does not satisfy the following transitivity implication: $\pi_i - \pi_k \geq \pi_j - \pi_k \geq 0 \implies D_{ik}^* \geq D_{jk}^*$. Therefore, in the current paper, we only assumed that D^* satisfies the local equivalence in Assumption 4.1, ensuring that small D_{ij}^* are bounded distortions of small $|\pi_i - \pi_j|/n$. This weak connection between D^* and π forced us to develop a more sophisticated local procedure to recover π .

On the negative side, the time complexity of our procedure is only bounded by $O(n^5)$, where n is the sample size of the observation $A \in \mathbb{R}^{n \times n}$. This high complexity comes from the heavy data-splitting that we used in `LocalRefineWS` (appendix H). We recall that the data-splitting was only made to facilitate the theoretical analysis of the estimator. In practice, it could be better to run the algorithm without data-splitting, i.e. to use `LocalRefine` instead of `LocalRefineWS`. Doing so, the time complexity will reduce from $O(n^5)$ to $O(n^3)$. However, because of complex statistical dependencies in this version without data-splitting, it is difficult to prove theoretical guarantees for the output $\hat{\pi}_o$. We only provided guarantees when each step of this algorithm are taken separately, and thus share no statistical dependencies.

The choice of tuning parameters $\delta_1, \delta_2, \delta_3$ is problematic, in general, since it depends on unknown constants, such as the parameters α, β of the class $\mathcal{BL}[\alpha, \beta]$ of bi-Lipschitz matrices. However, for a sufficiently large n , we proved that the choice of inputs becomes easier, and it is possible to choose $\delta_1, \delta_2, \delta_3$ as a function of n , independently of other parameters.

Acknowledgments

The work of C. Giraud has been partially supported by grant ANR-19-CHIA-0021-01 (BiSCottE, ANR) and ANR-21-CE23-0035 (ASCAI, ANR). The work of N. Verzelen has been partially supported by ANR-21-CE23-0035 (ASCAI, ANR).

A Models in the literature

Example 1: Monotone Toeplitz Matrix (a.k.a. Toeplitz R-matrix): Given a vector $(\theta_0, \theta_1, \dots, \theta_{n-1})$, a symmetric Toeplitz matrix F is defined by $F_{ij} = \theta_{|i-j|}$, for $i, j \in [n]$. If the vector is monotone $\theta_0 > \theta_1 > \dots > \theta_{n-1}$, the matrix F is called a monotone Toeplitz matrix. Thus, F is a special instance of R-matrix (1). This model has been recently studied in noisy seriation [Cai and Ma, 2022]. These matrices are equivalent to geometric 1D latent space model in Example 3.

Example 2: 1D latent space models: As probabilistic tools, latent space models are widely-used to study pairwise information data like networks [Hoff et al., 2002]. In 1D latent space models, the similarity matrix A is assumed to be sampled as follows. The distribution is parametrized by a 1D metric space (\mathcal{X}, d) , some (possibly random) latent positions $x_1, \dots, x_n \in \mathcal{X}$ and an similarity function $f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. Then, conditionally on x_1, \dots, x_n , the upper-diagonal entries A_{ij} of the similarity matrix are sampled independently, with conditional mean $\mathbb{E}[A_{ij} | x_1, \dots, x_n] = f(x_i, x_j)$.

This latent space formulation encompasses many classical models, such as graphons and f -Random Graphs [Diaconis and Janson, 2007, Lovász, 2012] and random geometric graphs [Penrose, 2003, Diaz et al., 2020, De Castro et al., 2017]. It also encompasses R-matrices (1), and monotone Toeplitz matrices (Example 1). To see that, take the latent space $\mathcal{X} = [0, 1]$, the latent positions $x_i = \pi_i/n$, and a similarity function f fulfilling $f(x_i, x_j) = F_{\pi_i, \pi_j}$.

Seriation in the 1D latent space models was considered in [Giraud et al., 2021] (with sub-Gaussian noise and real valued observations $A_{ij} \in \mathbb{R}$), and in [Janssen and Smith, 2020, Natick and Smith, 2021] (with f -random graphs, i.e., with binary observations $A_{ij} \in \{0, 1\}$ and a

uniform sample $x_1, \dots, x_n \sim U_{[0,1]}$ of the latent space $\mathcal{X} = [0, 1]$). The objective of recovering π is equivalent to re-ordering the latent positions x_1, \dots, x_n in the 1D space \mathcal{X} .

Example 3: Geometric 1D latent space model. As a special case of the 1D latent space model (Example 2), the geometric 1D latent space model is characterized by a similarity function of the form $f(x_i, x_j) = g(m(x_i, x_j))$ for a real non-increasing function $g : [0, \infty) \rightarrow \mathbb{R}$ and a metric m on \mathcal{X} . Thus, the similarity $f(x_i, x_j)$ depends on the latent positions x_i, x_j only via their distance $m(x_i, x_j)$. When the observations $A_{ij} \in \{0, 1\}$ are binary, this model is an instance of the popular random geometric graph [Penrose, 2003, Diaz et al., 2020, De Castro et al., 2017].

The geometric 1D latent space model is equivalent to the monotone Toeplitz matrices (Example 2). Indeed, for a vector $\theta_0 > \dots > \theta_{n-1}$ defining the monotone Toeplitz matrix $F_{ij} = \theta_{|i-j|}$, one can set a geometric 1D latent space model as follows. Take the latent space $\mathcal{X} = [0, 1]$ endowed with the metric $m(x, y) = |x - y|$, and the positions $x_i = \pi_i/n$, and a similarity function $f = g \circ m$ where g fulfills $g(t/n) = \theta_t$ for all $t \in \{0, \dots, n-1\}$. Thus, (for deterministic x_i 's) the two models lead to the same mean $\mathbb{E} A_{ij} = f(x_i, x_j) = g(|x_i - x_j|) = \theta_{|\pi_i - \pi_j|} = F_{\pi_i \pi_j}$.

By contrast, the latent bi-Lipschitz matrices (studied in the current paper) are not (fully) determined by some distances $m(x_i, x_j)$. Their general similarities $F_{\pi_i, \pi_j} = f(x_i, x_j)$ depend on the positions x_i, x_j , in such a way that, they can take different values $f(x_i, x_j) \neq f(x_k, x_l)$ even on points at a same distance $m(x_i, x_j) = m(x_k, x_l)$.

For statistical seriation in geometric 1D latent space models, see [Janssen and Smith, 2020], and [Giraud et al., 2021, section 4] and [Natik and Smith, 2021].

B Description of sub-algorithms

To complete the description of our procedure, we present the sub-algorithms `DistanceEstimation`, `LocalBisection` and `Orientation` in appendix B.1, B.2, B.3, respectively.

B.1 Distance estimation

We give the construction of the distance estimator \widehat{D} . The l_2 -type distance D^* in (11) is associated with a structure of inner product: Given vectors V and \tilde{V} in \mathbb{R}^n , we write their (normalized) inner product $\langle V, \tilde{V} \rangle_n = n^{-1} \sum_{\ell=1}^n V_\ell \tilde{V}_\ell$.

Algorithm 4 DistanceEstimation

Require: $A = [A_{ij}]_{i,j \in [n]}$ data matrix.

Ensure: \widehat{D} an $n \times n$ matrix.

- 1: **for** $i = 1, \dots, n$ **do**
 - 2: $\widehat{m}_i = \arg \min_{j \in [n]: j \neq i} \max_{k \in [n]: k \neq i, j} |\langle A_k, A_i - A_j \rangle_n|$.
 - 3: **end for**
 - 4: **for** $i, j = 1, \dots, n, i < j$ **do**
 - 5: $\widehat{D}_{ij} = \langle A_i, A_{\widehat{m}_i} \rangle_n + \langle A_j, A_{\widehat{m}_j} \rangle_n - 2\langle A_i, A_j \rangle_n$.
 - 6: $\widehat{D}_{ji} = \widehat{D}_{ij}$.
 - 7: **end for**
-

Let us explain the construction of the distance estimator \widehat{D} . Denoting by F_i the i^{th} column of the signal matrix F , we have the following decomposition of the distance D^* :

$$\forall i, j \in [n], i < j: \quad (D_{ij}^*)^2 = \langle F_{\pi_i}, F_{\pi_i} \rangle_n + \langle F_{\pi_j}, F_{\pi_j} \rangle_n - 2\langle F_{\pi_i}, F_{\pi_j} \rangle_n. \quad (22)$$

We estimate separately the crossed term and the two quadratic terms of (22).

◦ *Crossed-term*: Denoting by A_i the i^{th} column of the data matrix A , we observe that $\langle A_i, A_j \rangle_n$ is a sum of $(n-2)$ i.i.d. random variables (since $A_{ii} = A_{jj} = 0$). The $n-2$ random variables $\{A_{ik}A_{jk} : k \in [n] \text{ and } k \neq i, j\}$ are independent with the same mean

$$\mathbb{E}[A_{ik}A_{jk}] = F_{\pi_i\pi_k}F_{\pi_j\pi_k}, \quad (23)$$

where the expectation \mathbb{E} is taken over the data distribution $\mathbb{P}_{(F,\pi)}$. It is therefore possible to use standard concentration bounds to prove that, with high probability, $\langle A_i, A_j \rangle_n$ is close to its mean $\langle F_{\pi_i}, F_{\pi_j} \rangle_n$; see Lemma E.2. The inner product $\langle A_i, A_j \rangle_n$ (between two different columns) is thus a consistent estimator of the crossed term $\langle F_{\pi_i}, F_{\pi_j} \rangle_n$ in (22).

◦ *Quadratic-term*: We cannot proceed in the same way for the quadratic term $\langle F_{\pi_i}, F_{\pi_i} \rangle_n$ in (22), since it would lead to an inconsistent estimation. Indeed, we have

$$\mathbb{E}[A_{ik}^2] = F_{\pi_i\pi_k}^2 + \sigma^2 \mathbb{E}[E_{ik}^2], \quad (24)$$

and hence the inner product $\langle A_i, A_i \rangle_n$ between the same column is an inconsistent estimator of the quadratic term $\langle F_{\pi_i}, F_{\pi_i} \rangle_n$. It is possible to work around this issue via the following nearest neighbor approximation, which replaces the quadratic term by a crossed term, so as to be back to the unbiased case (23). Specifically, the approximation consists in replacing an object i by its nearest neighbor with respect to the distance D^* . Let $m_i \in \{1, \dots, n\}$, $m_i \neq i$, denote a nearest neighbor of i according to the distance D^* , that is $m_i \in \operatorname{argmin}_{t:t \neq i} D_{it}^*$. Then, we have the following approximation:

$$\begin{aligned} |\langle F_{\pi_i}, F_{\pi_i} \rangle_n - \langle F_{\pi_i}, F_{\pi_{m_i}} \rangle_n| &= |\langle F_{\pi_i}, F_{\pi_i} - F_{\pi_{m_i}} \rangle_n| \\ &\leq \frac{\|F_{\pi_i}\|_2}{\sqrt{n}} D_{im_i}^* \leq D_{im_i}^* \end{aligned} \quad (25)$$

using Cauchy-Schwarz inequality in the first inequality, and $F_{\pi_i} \in [0, 1]^n$ in the last one. The nearest neighbor approximation (25) thus yields a bias in our estimation procedure, which is equal to the distance $D_{im_i}^*$ between i and its nearest neighbor m_i . Since m_i is unknown, we still need to estimate m_i , and define an estimator \hat{m}_i . This step is performed in line 2 of Algorithm 4. Comments on it can be found in [Issartel, 2021, section 4.1]. We finally obtain an estimator $\langle A_i, A_{\hat{m}_i} \rangle_n$ of the quadratic term $\langle F_{\pi_i}, F_{\pi_i} \rangle_n$ in (22).

Putting things together, we get the estimator in line 5 of Algorithm 4. More information on this distance estimator can be found in appendix E.

B.2 Local Bisection

`LocalBisection` was described in plain words below Algorithm 2. Here we give its pseudo code. The inputs are an index $i \in [n]$, a symmetric matrix $D \in \mathbb{R}^{n \times n}$, and tuning parameters $\delta_1, \delta_2, \delta_3 > 0$. It performs a rough bisection of the set $[n]$ in i , so as to output two sets

$G_i^{(1)}, G_i^{(2)}$, one on each side of i (with respect to the ordering π). We use the notation $a \vee b = \max(a, b)$.

Algorithm 5 LocalBisection

Require: $(i, D, \delta_1, \delta_2, \delta_3)$

Ensure: $G_i^{(1)}, G_i^{(2)}$

- 1: Build a graph \mathcal{G}_i with node set $[n]$, by linking all $k, \ell \in [n]$, $k, \ell \neq i$ such that

$$D_{k\ell} \leq \delta_1 \quad \text{and} \quad D_{ik} \vee D_{i\ell} \geq \delta_2 \quad . \quad (26)$$

- 2: Collect only the connected components of \mathcal{G}_i that include (any) $k \in [n]$ such that

$$D_{ki} \geq \delta_3 \quad . \quad (27)$$

- 3: Denote by $G_i^{(1)}, G_i^{(2)}$ the two connected components with the largest cardinal numbers.
-

LocalBisection builds a graph \mathcal{G}_i according to the rule (26). Among all connected components of \mathcal{G}_i , only those satisfying (27) are collected, and then only the two largest of these components are output by the algorithm. We have the following interpretation, in the ideal scenario where the inputs D_{ij} are equal to the latent distances $|\pi_i - \pi_j|/n$. The rule (26) connects two nodes k, ℓ , if they are close to each other ($D_{k\ell} \leq \delta_1$), but one of them is far from i ($D_{ik} \vee D_{i\ell} \geq \delta_2$). Thus, such a connected component should be either on the right side of i , or on the left side of i . The second rule (27) selects the connected components of \mathcal{G}_i that contains (at least) a distant object from i ($D_{ki} \geq \delta_3$). As will be shown in appendix D.3, at most two connected components satisfy the condition (27), hence the outputs $G_i^{(1)}, G_i^{(2)}$ are necessarily these components, and the line 3 of LocalBisection is theoretically superfluous. We will see that $G_i^{(1)}, G_i^{(2)}$ are each on a different side of i .

B.3 Orientation

Orientation takes as inputs n pairs $(G_1^{(1)}, G_1^{(2)}, \dots, G_n^{(1)}, G_n^{(2)})$ of sets $G_i^{(1)}, G_i^{(2)} \subset [n]$, and outputs a new arrangement $[L_i, R_i]_{i \in [n]}$ of these pairs, i.e. $L_i, R_i \in \{G_i^{(1)}, G_i^{(2)}\}$ and $L_i \neq R_i$ (when $G_i^{(1)} \neq \emptyset, G_i^{(2)} \neq \emptyset$). The objective is to give a common orientation for the n new pairs L_i, R_i , $i = 1, \dots, n$, e.g., all L_i are on the left side while all R_i are on the right side of i . To do so, we choose an arbitrary direction L_{i^*}, R_{i^*} for an index i^* , and then use this direction as a reference for all other indices $i \in [n]$, $i \neq i^*$. We will show in appendix D.3 that Orientation gives n pairs L_i, R_i , $i = 1, \dots, n$, sharing a same direction.

Algorithm 6 Orientation

Require: $(G_i^{(1)}, G_i^{(2)})_{i \in [n]}$.**Ensure:** $(L, R) := (L_i, R_i)_{i \in [n]}$

```
1: Let  $V_{\neq \emptyset} = \{i \in [n] : G_i^{(1)} \neq \emptyset \text{ and } G_i^{(2)} \neq \emptyset\}$  and  $V'_{\neq \emptyset} = \{i \in [n] : G_i^{(1)} \neq \emptyset \text{ or } G_i^{(2)} \neq \emptyset\}$ .
2: if  $V_{\neq \emptyset} = \emptyset$  then
3:    $L_i = R_i = 0$  for all  $i \in [n]$ ; stop algorithm.
4: else
5:   let  $i^* \in V_{\neq \emptyset}$ , set  $L_{i^*} = G_{i^*}^{(1)}$  and  $R_{i^*} = G_{i^*}^{(2)}$ .   %Picking a reference direction  $L_{i^*}, R_{i^*}$ 
6: end if
7: for  $i \notin V_{\neq \emptyset} \cup V'_{\neq \emptyset}$  do
8:    $L_i = R_i = \emptyset$ .
9: end for
10: for  $i \in V_{\neq \emptyset}, i \neq i^*$  do
11:   if  $G_i^{(u)} \cap L_{i^*} = \emptyset$  for an  $u \in [2]$  then
12:      $R_i = G_i^{(u)}$  and  $L_i = G_i^{(v)}$  for  $v \in [2], v \neq u$    %Setting  $L_i, R_i$  for  $i \neq i^*, i \in V_{\neq \emptyset}$ 
13:   else if  $G_i^{(u)} \cap R_{i^*} = \emptyset$  for an  $u \in [2]$  then
14:      $L_i = G_i^{(u)}$  and  $R_i = G_i^{(v)}$  for  $v \in [2], v \neq u$ 
15:   end if
16: end for
17: for  $i \in V'_{\neq \emptyset}$  do
18:   Rearrange the notations  $G_i^{(1)}$  and  $G_i^{(2)}$  to obtain  $G_i^{(1)} \neq \emptyset$ .
19: end for
20: for  $i \in V'_{\neq \emptyset}$  do
21:   for  $k \in V_{\neq \emptyset}$  such that  $G_i^{(1)} \cap L_k \neq \emptyset$  and  $G_i^{(1)} \cap R_k \neq \emptyset$  do
22:     if  $i \in L_k$  then
23:        $L_i = \emptyset$  and  $R_i = G_i^{(1)}$ ; end for.   %Setting  $L_i, R_i$  for  $i \in V'_{\neq \emptyset}$ 
24:     else if  $i \in R_k$  then
25:        $R_i = \emptyset$  and  $L_i = G_i^{(1)}$ ; end for.
26:     end if
27:   end for
28: end for
```

In line 5, we take a reference index $i^* \in V_{\neq \emptyset}$ that has two non-empty sets, and we arbitrary choose its direction L_{i^*}, R_{i^*} . To align the sets $G_i^{(1)}, G_i^{(2)}$ of any i with the direction of L_{i^*}, R_{i^*} , we look at the values of (the four) intersections between these sets. The exact procedure depends on whether i has empty sets (line 7-8), or two non-empty sets (line 10-16), or one non-empty set (line 20-28). Since the i 's with (exactly) one non-empty set are more difficult to align, for them we may need to replace the initial reference sets L_{i^*}, R_{i^*} by another pair $L_k, R_k, k \in V_{\neq \emptyset}$ which provides more information on i 's set orientation (lines 21-27).

In line 11 and 13, the algorithm tests the emptiness of four intersections at most (for $u \in [2]$), and it stops as soon as one intersection is found empty. The testing order for $u \in [2]$ does not matter in our analysis.

C Proof of Theorem 2.1

Proof. (Theorem 2.1 follows from Theorem 4.1) The next two lemmas show that the bi-Lipschitz Assumption 2.1 implies the local distance equivalence Assumption 4.1, and the separated cumulative similarities Assumption 4.2.

Lemma C.1. *If F belongs to the class $\mathcal{BL}[\alpha, \beta]$ of bi-lipschitz matrices (as in Assumption 2.1), then, for $n \geq 8$, the distance matrix D^* is in $\mathcal{LDE}[\alpha/2, \beta, 0, r]$ (as in Assumption 4.1) for any $r \in (0, 1/4)$.*

Lemma C.1 states that, when F is a bi-Lipschitz matrix of $\mathcal{BL}[\alpha, \beta]$, the distance matrix D^* belongs to the class $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ for the parameters $\tilde{\beta} = \beta$ and $\tilde{\alpha} = \alpha/2$, thus conserving the Lipschitz upper constant β intact, and the lower constant α up to a factor $1/2$. The proof of Lemma C.1 is in appendix D.1.

Lemma C.2. *If $F \in \mathcal{BL}[\alpha, \beta]$ as in Assumption 2.1, then Assumption 4.2 holds for $\gamma = \alpha/4$, any $r' \in (0, 1/5)$ and any $r'' \in [0, 1/10)$, when $n \geq 20$.*

Lemma C.2 states that, when $F \in \mathcal{BL}[\alpha, \beta]$, Assumption 4.2 holds for the constants $\gamma = \alpha/4$, any $r' \in (0, 1/5)$ and any $r'' \in [0, 1/10)$, hence a conservation of the Lipschitz lower constant α up to a factor $1/4$. The proof of Lemma C.2 is in appendix D.1.

Thus, Theorem 2.1 follows from Theorem 4.1. \square

In order to prove Theorem 4.1, the rest of the section is organized as follows. We give in appendix C.1 the general versions of Theorem 2.1 and 4.1 where α , β , and σ may depend on n , but the choices of the tuning parameters $\delta_1, \delta_2, \delta_3$ are more intricate. (This appendix C.1 may be skipped if the reader is only interested in the proof of Theorem 4.1.) Next, we present guarantees for the sub-algorithms of SALB in the following order: the distance estimation is in appendix C.2, the first and second estimators of $H^{(\pi)}$ in appendix C.3 and C.4, respectively, and the transition from comparison matrix to permutation in appendix C.5. Finally, the proof of Theorem 4.1 (and of its general version, Theorem C.4) is in appendix C.6.

C.1 Choice of tuning parameters $\delta_1, \delta_2, \delta_3$ and general results

For Theorem 2.1: To simplify, Theorem 2.1 was stated in a setting where α , β , and σ are considered as constants whereas n is large, and where the tuning parameters $\delta_1, \delta_2, \delta_3$ of Algorithm 1 was set to specific values $\delta_1 = n^{-1/5}$, $\delta_2 = \log(n)\delta_1$ and $\delta_3 = \log(n)\delta_2$. However, SALB achieves the same convergence rate even in settings where the quantities α, β, σ depend on n , but the choices of the tuning parameters $\delta_1, \delta_2, \delta_3$ are more intricate. The choices of $\delta_1, \delta_2, \delta_3$ may be summarized by the following conditions:

$$C_{\beta\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \leq \delta_1, \quad C_{\alpha\beta} \delta_1 \leq \delta_2 \leq C_\alpha, \quad C'_{\alpha\beta} \delta_2 \leq \delta_3 \leq C''_{\alpha\beta}, \quad (28)$$

where $n \geq 8$, and $C_{\beta\sigma}, C_{\alpha\beta}, C_\alpha, C'_{\alpha\beta}, C''_{\alpha\beta}$ are constants only depending on (some of) the problem parameters α, β, σ . The exact conditions (with explicit constants) for the choices of $\delta_1, \delta_2, \delta_3$ can be found in (112-113).

The next result generalizes Theorem 2.1 to a setting where α, β, σ may depend on n , and the $\delta_1, \delta_2, \delta_3$ are chosen in the admissible range (28).

Theorem C.3. *For $n \geq 8$ and $\alpha, \beta, \sigma > 0$, there exist positive constants $C_{\beta\sigma}, C_{\alpha\beta}, C_\alpha, C'_{\alpha\beta}, C''_{\alpha\beta}$ such that the following holds for any $\delta_1, \delta_2, \delta_3$ fulfilling (28) and for any $F \in \mathcal{BL}[\alpha, \beta]$. With probability $1 - 9/n^2$, the permutation $\hat{\pi}_o$ output by SALB satisfies*

$$L_{\max}(\hat{\pi}_o, \pi) \leq C \frac{\sigma}{\alpha} \sqrt{\frac{\log(n)}{n}}, \quad (29)$$

where $C > 0$ is a numerical constant.

The choices of $\delta_1, \delta_2, \delta_3$ enforced by (28) may seem purely theoretical since they depend on unknown problem parameters α, β, σ (via the constants $C_{\beta, \sigma}, \dots$). In fact, (28) shows that any $\delta_1, \delta_2, \delta_3$ satisfying the following conditions, as $n \rightarrow \infty$,

$$\frac{\delta_1}{(\log(n)/n)^{1/4}} \rightarrow \infty, \quad \frac{\delta_2}{\delta_1} \rightarrow \infty, \quad \frac{\delta_3}{\delta_2} \rightarrow \infty, \quad \delta_3 \rightarrow 0, \quad (30)$$

will be a solution of (28) as soon as n is larger than some constant $C_{\alpha\beta\sigma}$. In particular, our choice of tuning parameter in Theorem 2.1, which was $\delta_1 = n^{-1/5}$, $\delta_2 = \log(n)\delta_1$ and $\delta_3 = \log(n)\delta_2$, satisfies (30). Thus, Theorem 2.1 follows from Theorem C.3. \square

◦ *For Theorem 4.1:* The same goes as for Theorem 2.1, except that the constants in (28) have new dependencies. More precisely, the dependencies in the bi-Lipschitz constants α and β are replaced by dependencies in the distance equivalence constants $\tilde{\alpha}$ and $\tilde{\beta}$, respectively, and there are new dependencies in r, r' and r'' . This gives

$$C_{\tilde{\beta}\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \leq \delta_1, \quad C_{\tilde{\alpha}\tilde{\beta}} \delta_1 \leq \delta_2 \leq C_{\tilde{\alpha}r'r''}, \quad C'_{\tilde{\alpha}\tilde{\beta}} \delta_2 \leq \delta_3 \leq C_{\tilde{\alpha}\tilde{\beta}r'r''}, \quad (31)$$

where $n \geq (1/2r) \vee 8$, and where $C_{\tilde{\beta}\sigma}, C_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}r'r''}, C'_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}\tilde{\beta}r'r''}$ are constants only depending on (some of) the problem parameters $\tilde{\alpha}, \tilde{\beta}, r, r', r'', \sigma$. The exact conditions (with explicit constants) for the choices of $\delta_1, \delta_2, \delta_3$ can be found in (115), taking $\omega = \omega_n$ and $\tilde{\rho} = \rho$ in (115). Note that the constant γ of Assumption 4.2 is not involved in the choice of tuning parameter. The next result generalizes Theorem 4.1 to a setting where $\tilde{\alpha}, \tilde{\beta}, \sigma, r, r', r''$ may depend on n , but the tuning parameters $\delta_1, \delta_2, \delta_3$ must fulfill (31).

Theorem C.4. *For $n \geq 8$ and $\tilde{\alpha}, \tilde{\beta}, \sigma, r, r', r'' >$, there exist positive constants $C_{\tilde{\beta}\sigma}, C_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}r'r''}, C'_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}\tilde{\beta}r'r''}$ such that the following holds for any $\delta_1, \delta_2, \delta_3$ fulfilling (31) and for any $F \in \mathcal{SCA}[\gamma, r', r'']$ such that $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$. With probability $1 - 9/n^2$, the permutation $\hat{\pi}_o$ output by SALB satisfies*

$$L_{\max}(\hat{\pi}_o, \pi) \leq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}},$$

where $C > 0$ is a numerical constant.

Theorem 4.1 follows from Theorem C.4 (in the same way that Theorem 2.1 followed from Theorem C.3 above). \square

We proved that Theorem 2.1 followed from Theorem 4.1 (at the beginning of appendix C), and it goes the same for their generalizations: Theorem C.3 follows from Theorem C.4. Thus, Theorem C.4 implies all the results Theorem 2.1, C.3 and 4.1.

C.2 Error of the distance estimator

The next proposition gives a local upper bound on the estimation error of \widehat{D} , that is, on the errors $|\widehat{D}_{ij} - D_{ij}^*|$ for all i, j within a distance r .

Proposition C.5. *For any $8 \leq n$, and $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r$, the following holds for any $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ (as defined in Assumption 4.1). With probability $1 - 1/n^4$, the matrix \widehat{D} is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega_n, r]$ for*

$$\omega_n = C \left(\sqrt{\frac{\tilde{\beta}}{n}} + \sqrt{(\sigma + 1)\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \right) \quad (32)$$

where $C > 0$ is a numerical constant.

Thus, for all i, j within a distance r , the estimate \widehat{D}_{ij} is equivalent to the distance $|\pi_i - \pi_j|/n$, up to factors $\tilde{\alpha}$ and $\tilde{\beta}$, and to the additive error ω_n . The proof in appendix D.2 shows that ω_n upper bounds the (local) estimation errors of distances D_{ij}^* .

C.3 Error of the first estimator of $H^{(\pi)}$

Since our intermediate objective is to estimate the matrix $H^{(\pi)}$, we introduce for convenience the notion of error for a comparison matrix H . We say that a comparison matrix H has an error smaller than ν , if it satisfies the following for an $s \in \{\pm\}$,

$$sH_{ij} = H_{ij}^{(\pi)} \quad \text{for all } i, j \text{ such that } \frac{|\pi_i - \pi_j|}{n} \geq \nu. \quad (33)$$

In words, H matches $H^{(\pi)}$ on all pairs i, j whose positions π_i, π_j in the ordering are at distance greater than ν . The sign ($s = \pm$) comes from the fact that π is identifiable up to a reverse of the ordering (section 1.1).

Recall that `AgregLocalBisection` takes as inputs a matrix D and tuning parameters $\delta_1, \delta_2, \delta_3$, and it outputs a comparison matrix H . Proposition C.6 states that, if the inputs D_{ij} are locally equivalent to the distances $|\pi_i - \pi_j|/n$ as in Assumption 4.1, then the output H coincides with the true comparison matrix $H^{(\pi)}$ on all entries (i, j) such that i, j are at a distance greater than some constant ρ . The proof is in appendix D.3.

Proposition C.6. *For any $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r$ and $0 \leq \omega$, the following holds for any $D \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ (as defined in Assumption 4.1) and any $\delta_1, \delta_2, \delta_3$ fulfilling (114). There exists $s \in \{\pm\}$ such that, the output H of `AgregLocalBisection` satisfies*

$$H_{ij} = sH_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } H_{ij} \neq 0 \quad \text{or} \quad \frac{|\pi_i - \pi_j|}{n} \geq \rho,$$

for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$

A remarkable property of H is to be correct on its support (i.e., for all entries $H_{ij} \neq 0$), and thus, we know which part of the matrix H is trustworthy. In addition, H is correct for all i, j that are at distance $|\pi_i - \pi_j|/n$ greater than ρ . This lower bound $\rho = (\delta_2 + \omega)/\tilde{\alpha}$ depends on some parameters of the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ class of the input D , namely, the additive error ω and the contraction factor $\tilde{\alpha}$ of the local distance equivalence in Assumption 4.1. One can see that the lower bound ρ deteriorates (i.e. becomes larger) when the ratio $\omega/\tilde{\alpha}$ gets bigger, i.e., when the minimum signal $\tilde{\alpha}|\pi_i - \pi_j|/n - \omega \leq D_{ij}$ (ensured by the \mathcal{LDE} assumption) becomes close to zero. This confirms the intuition that, when the inputs D_{ij} are less uninformative about the distances $|\pi_i - \pi_j|/n$, the performance of `AgregLocalBisection` deteriorates.

Although none of the usual properties of distances are assumed (e.g. transitivity, or triangle inequality), Proposition C.6 shows that our distance based method can partially recover the matrix of comparison $H^{(\pi)}$.

In Proposition C.6, we assumed the conditions (114) on $\delta_1, \delta_2, \delta_3$; these conditions can be summarized by (31) when $\omega = \omega_n$ for the ω_n defined in (32).

C.4 Error of the second estimator of $H^{(\pi)}$

Recall that `LocalRefine` takes as input a comparison matrix $H_1 \in \{-1, 0, 1\}^{n \times n}$ and outputs another comparison matrix H . When the input H_1 is a good enough estimate of $H^{(\pi)}$ to fulfill (34), then, Proposition C.7 ensures that the output H improves on the accuracy of H_1 , recovering the remaining entries (i, j) up to up to a distance $(\sigma/\gamma)\sqrt{\frac{\log(n)}{n}}$ between i, j , as described in (35). In plain words, if H_1 is correct on its support and on all its entries (i, j) such that i, j are at distance greater than a constant $\tilde{\rho}$, then, H will successfully recover any comparison $H_{ij}^{(\pi)}$ left undetermined by H_1 such that i, j are at distance greater than $(\pi/\gamma)\sqrt{\log n/n}$.

Proposition C.7. *For any $0 < \gamma, r'$ and $0 \leq r''$, the following holds for any $F \in \mathcal{SCA}[\gamma, r', r'']$ (as defined in Assumption 4.2) and any $\tilde{\rho} \in [0, r' \wedge r'']$. If the the input H_1 of `LocalRefine` is deterministic or independent of the data A , with the following accuracy, for any $\epsilon \in \{\pm\}$,*

$$(H_1)_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } (H_1)_{ij} \neq 0 \text{ or } \frac{|\pi_i - \pi_j|}{n} \geq \tilde{\rho}, \quad (34)$$

then, with probability $1 - 4/n^3$, the output H of `LocalRefine` satisfies for all i, j ,

$$H_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{whenever } (H_1)_{ij} = 0 \text{ and } \frac{|\pi_i - \pi_j|}{n} \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}. \quad (35)$$

In addition, $H_{ij} = 0$ whenever $(H_1)_{ij} \neq 0$.

Among the comparisons $H_{ij}^{(\pi)}$ left undetermined by the input H_1 , `LocalRefine` correctly estimates the ones fulfilling (35), i.e., those associated with distances $|\pi_i - \pi_j|/n$ greater than $(C\sigma/\gamma)\sqrt{\log(n)/n}$. The distance $(\sigma/\gamma)\sqrt{\log(n)/n}$ shows that the performance of H is declining with the input σ , but is improving with the level of separation γ of cumulative similarities of $F \in \mathcal{SCA}[\gamma, r', r'']$.

The proof of Proposition C.7 is in Appendix D.4. The next proposition gives guarantees for `LocalrefineWS` performance; it looks essentially the same as Proposition C.7, up to technical details related to the data splitting in `LocalrefineWS`.

Proposition C.8. For any $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r, \gamma, r'$ and $0 \leq \omega, r'', \tilde{\rho}$, the following holds for any $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ and $D \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ (as defined in Assumption 4.1), any $F \in \text{SCA}[\gamma, r', r'']$ (as defined in Assumption 4.2) and any $\delta_1, \delta_2, \delta_3, \tilde{\rho}$ fulfilling (115). If the input H_1 of `LocalRefineWS` has the following accuracy, for any $\epsilon \in \{\pm\}$,

$$(H_1)_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } (H_1)_{ij} \neq 0 \text{ or } \frac{|\pi_i - \pi_j|}{n} \geq \tilde{\rho}, \quad (36)$$

then, with probability $1 - 8/n^2$, the output H satisfies

$$H_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{whenever } (H_1)_{ij} = 0 \text{ and } \frac{|\pi_i - \pi_j|}{n} \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}. \quad (37)$$

In addition, $H_{ij} = 0$ whenever $(H_1)_{ij} \neq 0$.

The proof of Proposition C.8 is relatively lengthy and technical, but the essence is a combination of the ideas from Proposition C.7, C.6 and C.5. The proof is delayed to appendix I.

C.5 Final step: from comparison matrix to permutation

Recall the last two lines of SALB (Algorithm 1): compute the scores $S = [\hat{H}_1 + \hat{H}_2]\mathbf{1}$; then build any permutation $\hat{\pi}_o$ by increasing values of S . This can be more generally stated as follows: Given a comparison matrix $H \in \{-1, 0, 1\}^{n \times n}$, define a permutation π_H by

$$\forall i \in [n]: S_i^H = \sum_{j \in [n]} H_{ij} \quad \text{and} \quad \pi_i^H = \#\{j \in [n]: S_j^H \leq S_i^H\}, \quad (38)$$

where we break ties arbitrarily, so that π^H is a permutation of $[n]$. We emphasize that any comparison matrix H has zero diagonal coefficients (i.e. $H_{ii} = 0$ for all i). In the definition (38), we can observe that $\pi^{H^{(\pi)}} = \pi$, where $H^{(\pi)}$ is the comparison matrix associated to π .

Recalling the notion of error (33) for a comparison matrix H , we want to prove that π^H is almost as accurate as H . Proposition C.9 states precisely this, that the error of π^H is bounded by that of H up to a factor 2. The proof is in Appendix D.5.

Proposition C.9. Let $\nu > 0$. If a comparison matrix H has an error less than ν as in (33), then the permutation π^H in (38) has an L_{\max} error less than 2ν , that is, $L_{\max}(\pi^H, \pi) \leq 2\nu$.

C.6 Proof of Theorem 4.1 and C.4

We run `SALB`($A, \delta_1, \delta_2, \delta_3, \sigma$) – Algorithm 1. Let \mathcal{E}_n be the event where \hat{D} – the output of `Distance Estimation`(A) – is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega_n, r]$ for the ω_n defined in (32). Proposition C.5 states that the probability of occurrence of the event \mathcal{E}_n is at least $1 - 1/n^4$.

We now want to apply Proposition C.6 and C.8 to obtain guarantees for $\hat{H}_1 + \hat{H}_2$ – the sum of the outputs of `AgreLocalBisection`($A, \hat{D}, \delta_1, \delta_2, \delta_3$) and `LocalRefineWS`($A, \hat{D}, \delta_1, \delta_2, \delta_3, \sigma$). Conditionally to the event \mathcal{E}_n , and if $\delta_1, \delta_2, \delta_3$ fulfill the conditions (114-115), Proposition C.6 and C.8 ensure that the following holds with probability at least $1 - 8/n^2$. For an $s \in \{\pm\}$, we have

$$(\hat{H}_1 + \hat{H}_2)_{ij} = s H_{ij}^{(\pi)} \quad \text{for all } i, j \text{ such that } |\pi_i - \pi_j| \geq C \frac{\sigma}{\gamma} \sqrt{n \log(n)}. \quad (39)$$

Parenthesis: [Let us make a parenthesis to unpack the step (39), placing us for a moment in the imaginary scenario where \widehat{H}_1 would be either deterministic or independent of the data A . In this scenario, we can run the simpler `LocalRefine` instead of `LocalRefineWS`. Then, since the input \widehat{D} of `AgregLocalBisection` is $\mathcal{LDE}[\widehat{\alpha}, \widehat{\beta}, \omega_n, r]$ on the event \mathcal{E} , Proposition C.6 guarantees that the output \widehat{H}_1 of `AgregLocalBisection` satisfies, for an $s \in \{\pm 1\}$,

$$(\widehat{H}_1)_{ij} = sH_{ij}^{(\pi)} \quad \text{wherever } (\widehat{H}_1)_{ij} \neq 0 \text{ or } \frac{|\pi_i - \pi_j|}{n} \geq \rho ,$$

where $\rho = (\delta_2 + \omega)/\widehat{\alpha}$. Then, if $\rho \leq r' \wedge r''$, we can take $\tilde{\rho} = \rho$ and $H = \widehat{H}_1$ in Proposition C.7. This yields, with probability $1 - 4/n^3$, the following accuracy for the output \widehat{H}_2 of `LocalRefine`:

$$(\widehat{H}_2)_{ij} = sH_{ij}^{(\pi)} \quad \text{wherever } (\widehat{H}_1)_{ij} = 0 \text{ and } \frac{|\pi_i - \pi_j|}{n} \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}} ,$$

and, in addition, $(\widehat{H}_2)_{ij} = 0$ on the support of \widehat{H}_1 (where $(\widehat{H}_1)_{ij} \neq 0$). Note that the aforementioned condition $\rho \leq r' \wedge r''$ is satisfied under the assumption that the tuning parameter $\delta_1, \delta_2, \delta_3$ fulfills the constraints (31). Putting the last two displays together, we obtain the accuracy (39), which closes the parenthesis.]

The output $\hat{\pi}_o$ of `SALB` can be written as $\hat{\pi}_o = \pi^{\widehat{H}_1 + \widehat{H}_2}$ using the formula (38) (taking $H = \widehat{H}_1 + \widehat{H}_2$ in the formula). Conditionally to the event (39), Proposition C.9 yields

$$L_{\max}(\hat{\pi}_o, \pi) \leq 2C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}} . \quad (40)$$

Taking a union bound over the events \mathcal{E}_n and (39), we conclude that the bound (40) holds with probability $1 - 9/n^2$ (using $8/n^2 + 1/n^4 \leq 9/n^2$). Thus, Theorem C.4 follows. Since Theorem 4.1 is a particular case of Theorem C.4 (see appendix C.1), Theorem 4.1 follows too.

D Proof of lemmas and propositions from Appendix C

D.1 Proofs of Lemma C.1 and C.2

Proof of Lemma C.1. We use the notation $y_j = j/n$ and $x_j = \pi_j/n$ for all $j \in [n]$. Let $r \in (0, 1/4)$, and $i, j \in [n]$ such that $|x_i - x_j| \leq r$. For $F \in \mathcal{BL}[\alpha, \beta]$ we have

$$D_{ij}^* = \left(\frac{1}{n} \sum_{\ell=1}^n [F_{\pi_i \ell} - F_{\pi_j \ell}]^2 \right)^{1/2} \leq \beta \frac{|\pi_i - \pi_j|}{n} .$$

This gives the upper bound in the $\mathcal{LDE}[\alpha/2, \beta, 0, r]$ condition.

Since $n \geq 8$ and $|x_i - x_j| \leq r$ for $r < 1/4$, one can readily check that $x_i, x_j \in (1/4 + 1/n, 1]$, or $x_i, x_j \in [0, 3/4 - 1/n]$. We only study the case $x_i, x_j \in (1/4 + 1/n, 1]$, the other case being symmetric. For $F \in \mathcal{BL}[\alpha, \beta]$ as in Assumption 2.1, we have

$$D_{ij}^* \geq \left(\frac{1}{n} \sum_{\ell: y_\ell \leq \frac{1}{4} + \frac{1}{n}} [F_{\pi_i, \ell} - F_{\pi_j, \ell}]^2 \right)^{1/2} \geq \alpha \frac{|\pi_i - \pi_j|}{n} \left(\frac{1}{n} \#\{\ell : y_\ell \leq \frac{1}{4} + \frac{1}{n}\} \right)^{1/2} .$$

We have $\#\{\ell : y_\ell \leq \frac{1}{4} + \frac{1}{n}\} \geq \frac{n}{4}$, so the lower bound in $\mathcal{LDE}[\alpha/2, \beta, 0, r]$ follows. \square

Proof of Lemma C.2. We use the notation $y_j = j/n$ and $x_j = \pi_j/n$ for all $j \in [n]$. Let $r' \in (0, 1/5)$, $r'' \in [0, 1/10)$ and $i, j \in [n]$, $i < j$, such that $|y_i - y_j| \leq r'$. Then, at least one of the two following inequalities holds: $y_i \geq \frac{1}{2} - \frac{r'}{2}$ or $y_j \leq \frac{1}{2} + \frac{r'}{2}$. We only study the case where $y_i \geq \frac{1}{2} - \frac{r'}{2}$. Then, we have

$$y_j > y_i > \frac{4}{10} > r'' + \frac{3}{10} .$$

For $n \geq 20$, it follows that

$$y_i - r'' > \frac{3}{10} \geq \frac{1}{4} + \frac{1}{n} .$$

This gives $\{\ell : y_\ell \leq \frac{1}{4} + \frac{1}{n}\} \subset \{\ell : y_\ell < y_i - r''\}$ and then $\#\{\ell : y_\ell < y_i - r''\} \geq n/4$. Combining with the assumption $F \in \mathcal{BL}[\alpha, \beta]$, we obtain

$$\sum_{\ell: y_\ell < y_i - r''} F_{i\ell} - F_{j\ell} \geq \alpha \frac{|i-j|}{n} \#\{\ell : y_\ell < y_i - r''\} \geq \frac{\alpha}{4} |i-j| .$$

Thus, Assumption 4.2 holds for $\gamma = \alpha/4$. \square

D.2 Proof of Proposition C.5

It is sufficient to show that, with high probability, the estimation error of \widehat{D} is locally bounded by ω_n , that is, $|(\widehat{D})_{ij} - D_{ij}| < \omega_n$ for all close i, j . Lemma D.1 makes this claim clear. Then, on this high probability event, the estimator \widehat{D} inherits the \mathcal{LDE} properties of the distance matrix D^* , up to an additive error ω_n . Proposition C.5 is proved. \square

Lemma D.1. *For any $4 \leq n$, and $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r$, the following holds for any $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ (as defined in Assumption 4.1). With probability $1 - 1/n^4$, we have*

$$|(\widehat{D})_{ij} - D_{ij}^*| < \omega_n \quad \text{for all } i, j \in [n] \text{ such that } \frac{|\pi_i - \pi_j|}{n} \wedge D_{ij}^* \leq r \quad (41)$$

where

$$\omega_n = C \left(\sqrt{\frac{\tilde{\beta}}{n}} + \sqrt{(\sigma + 1)\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \right)$$

and C is a numerical constant.

When the true distance matrix D^* is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ (as in Assumption 4.1), Lemma D.1 gives, with high probability, a uniform control on the estimation errors $|(\widehat{D})_{ij} - D_{ij}^*|$, for all i, j at distance less than r . The error bound ω_n is increasing with σ (the noise level bound), and with $\tilde{\beta}$ (the upper bound on the distortions of $(D^*)_{ij}$ with respect to $|\pi_i - \pi_j|/n$). In particular, when $\tilde{\beta}, \sigma$ are fixed constants, the bound ω_n goes to zero (as $n \rightarrow \infty$) and thus \widehat{D} is locally a consistent estimator of D^* . The proof of Lemma D.1 is in appendix E.

D.3 Proof of Proposition C.6

We analyze separately the three steps of `AgregLocalBisection`. More precisely, the performance of the first step `LocalBisection` is given by Lemma D.3, the second step `Orientation` by Lemma D.4, and the last step (lines 6-12 of `AgregLocalBisection` which set the values of a comparison matrix H) by Lemma D.5. Proposition C.6 will then follow directly from these three lemmas.

D.3.1 The first step: `LocalBisection`

We start by giving a rationale for `LocalBisection`: Lemma D.2 gives theoretical guarantees in the ideal situation where the inputs D_{ij} are equal to the latent distances $|\pi_i - \pi_j|/n$. More precisely, it states that: (1) at most two connected components of the graph \mathcal{G}_i satisfy the condition (27) (hence the line 3 of `LocalBisection` is theoretically unnecessary), and (2) each of these components is on a (different) side of i (in the ordering π).

Before stating Lemma D.2, we need some notations. For any $i \in [n]$, denote by L_i^* (resp. R_i^*) the set of objects k such that $\pi_k < \pi_i$ (resp. $\pi_k > \pi_i$). For any $\lambda > 0$, denote by $sL_i^*(\lambda)$ (resp. $sR_i^*(\lambda)$) the set of object k such that $\pi_k \leq \pi_i - n\lambda$ (resp. $\pi_k \geq \pi_i + n\lambda$). Thus, $sL_i^*(\lambda)$ (resp. $sR_i^*(\lambda)$) is the subset of L_i^* (resp. R_i^*) reduced by the length λ . We now define two properties:

- **Property $\mathcal{P}_i(\lambda, G)$:** given any $\lambda > 0$ and any $G \subset [n]$, we have the inclusions

$$\emptyset \neq sL_i^*(\lambda) \subset G \subset L_i^* \quad \text{or} \quad \emptyset \neq sR_i^*(\lambda) \subset G \subset R_i^* . \quad (42)$$

- **Property $\mathcal{P}'_i(\lambda, G^{(1)}, G^{(2)})$:** given any $\lambda > 0$ and any $G^{(1)}, G^{(2)} \subset [n]$, we have

$$\emptyset \neq sL_i^*(\lambda) \subset G^{(u)} \subset L_i^* \quad \text{and} \quad \emptyset \neq sR_i^*(\lambda) \subset G^{(v)} \subset R_i^* , \quad (43)$$

for some $u, v \in [2], u \neq v$.

Lemma D.2 (ideal input D). *Let $i \in [n]$. If the inputs in `LocalBisection`($i, D, \delta_1, \delta_2, \delta_3$) are such that, $1/n \leq \delta_1 \leq \delta_2 \leq \delta_3 \leq 1/4$, and $D_{ij} = |\pi_i - \pi_j|/n$ for all $i, j \in [n]$, then:*

1. *at least one and at most two connected components of \mathcal{G}_i satisfy the condition (27). We write $G_i^{(1)}, G_i^{(2)}$ these outputs, with the convention $G_i^{(1)} \neq \emptyset$, while $G_i^{(2)}$ may be empty.*
2. *When $G_i^{(2)} = \emptyset$, the set $G_i^{(1)}$ is on one side of i , and includes all objects δ_2 away from i on this side (w.r.t. the distance $D_{ij} = |\pi_i - \pi_j|/n$). More precisely, property $\mathcal{P}_i(\delta_2, G_i^{(1)})$ in (42) is satisfied.*
3. *When $G_i^{(2)} \neq \emptyset$, the sets $G_i^{(1)}, G_i^{(2)}$ are on opposite sides of i , and include together all objects δ_2 away from i . More precisely, property $\mathcal{P}'_i(\delta_2, G_i^{(1)}, G_i^{(2)})$ in (43) is satisfied.*

Thus, the outputs $G_i^{(1)}, G_i^{(2)}$ are on different sides of i in the ordering π , one being included in the left set L_i^* , while the other is in the right set R_i^* . In addition, $G_i^{(1)}, G_i^{(2)}$ together contain all objects at distance δ_2 from i , one including $sL_i^*(\delta_2)$, the other $sR_i^*(\delta_2)$. Therefore, `LocalBisection` recovers left and right sets of i with an accuracy δ_2 . The interest of this lemma is to show the mechanics of `LocalBisection` in a simple situation (where we have ideal inputs $D_{ij} = |\pi_i - \pi_j|/n$). The proof of Lemma D.2 is in appendix F.1.

Lemma D.3 generalizes Lemma D.2 to the situation considered in this paper where the quantities $|\pi_i - \pi_j|/n$ are unknown, and our input D can deviate from these ideal quantities as much as permitted by Assumption 4.1.

Lemma D.3. *Let $i \in [n]$. If the inputs in `LocalBisection`($i, D, \delta_1, \delta_2, \delta_3$) are such that, D is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ as in Assumption 4.1, and $\delta_1, \delta_2, \delta_3$ fulfill the constraints*

$$\delta_1 \geq \omega + \frac{\tilde{\beta}}{n} , \quad \delta_2 > \omega + \frac{\tilde{\beta}}{\tilde{\alpha}}(\delta_1 + \omega) , \quad r \geq \frac{n^{-1} \vee (\delta_1 + \omega) \vee (\delta_2 + \omega)}{1 \wedge \tilde{\alpha}} , \quad (44)$$

$$n \geq 4 , \quad (r \wedge (\tilde{\alpha}/4)) - \omega \geq \delta_3 > \omega + \tilde{\beta}\rho ,$$

for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$. Then, the points 1,2,3 of Lemma D.2 hold for an accuracy ρ instead of δ_2 , that is:

1. point 1 is unchanged,
2. points 2 holds for $\mathcal{P}_i(\rho, G_i^{(1)})$ instead of $\mathcal{P}_i(\delta_2, G_i^{(1)})$,
3. points 3 holds for $\mathcal{P}'_i(\rho, G_i^{(1)}, G_i^{(2)})$ instead of $\mathcal{P}'_i(\delta_2, G_i^{(1)}, G_i^{(2)})$.

Thus, when the input D is in $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$, `LocalBisection` successfully recovers left and right sets with an accuracy $\rho = (\delta_2 + \omega)/\tilde{\alpha}$. The proof is in appendix F.2.

Remark D.1. For the inputs $\delta_3 = \delta_2 = \delta_1 = 1/n$ in Lemma D.2, we get a perfect recovery of the left and right groups L_i^*, R_i^* of i . This is not true anymore in the realistic scenario of Lemma D.3, where the recovery error $\rho = (\delta_2 + \omega)/\tilde{\alpha}$ is possibly large, in part because of the constraints (44) which prevent from choosing a small value for the tuning parameter δ_2 .

Remark D.2. The rule (27) of `LocalBisection` removes the connected components that do not have a node at distance greater than δ_3 . Since δ_3 must be sufficiently large to satisfy the constraints (44), the rule (27) may cause abusive deletions of graph components. As a consequence, we may (too) often have an empty output $G_i^{(2)}$ in the `LocalBisection`, resulting in a loss of information. To circumvent this issue, we use a double checking to set H in `AgregLocalBisection` (the first checking is in lines 7-8, and the second checking in lines 9-10).

D.3.2 The second step: Orientation

Lemma D.4 ensures that `Orientation` will rearrange the outputs $[G_i^{(1)}, G_i^{(2)}]_{i \in [n]}$ of `LocalBisection` to obtain n pairs $[L_i, R_i]_{i \in [n]}$ that share a same orientation.

Lemma D.4. *Assume that the inputs $[G_i^{(1)}, G_i^{(2)}]_{i \in [n]}$ of `Orientation` are equal to the outputs of `LocalBisection`. With no loss of generality, assume that the following orientation is taken in the line 5 of `Orientation`:*

$$\arg \min_{i \in L_i^* \cup R_i^*} \pi_i \in L_i^* . \quad (45)$$

Under the hypotheses of Lemma D.3, and the constraints $n^{-1} \leq \rho \leq 1/8$ and $\delta_3 + \omega \leq r \wedge (\tilde{\alpha}/8)$, the outputs L_i, R_i of `Orientation` satisfy, for all $i \in [n]$,

$$\begin{aligned} L_i \cup R_i &\neq \emptyset , \\ \emptyset \neq sL_i^*(\rho) &\subset L_i \subset L_i^* && \text{if } L_i \neq \emptyset , \\ \emptyset \neq sR_i^*(\rho) &\subset R_i \subset R_i^* && \text{if } R_i \neq \emptyset . \end{aligned} \quad (46)$$

By property (46), the outputs $[L_i, R_i]_{i \in [n]}$ of **Orientation** are such that, all sets L_i are on the left side while all sets R_i are on the right side of i . The proof of Lemma D.4 is in appendix F.3.

D.3.3 The last step: lines 6-12 of **AgregLocalBisection**

Having sets $[L_i, R_i]_{i \in [n]}$ satisfying (46) at our disposal, it would be natural to define a comparison matrix H as follows:

$$\forall i \in [n]: H_{ij} = 1 \text{ for } j \in L_i \quad \text{and} \quad H_{ij} = -1 \text{ for } j \in R_i, \quad (47)$$

leaving undetermined the other entries (i.e., $H_{ij} = 0$ for $j \notin L_i \cup R_i$). This gives correct estimates H_{ij} of true comparisons $H_{ij}^{(\pi)}$ for all $j \in L_i \cup R_i$. However, this single checking is not sufficient to prove good performances for H , and accordingly, the proper construction of H in lines 6-12 of **AgregLocalBisection** is slightly less direct, and uses a double checking instead of one. This double checking is discussed in Remark D.2 and D.3.

Lemma D.5. *Under the hypotheses of Lemma D.4, the output H of **AgregLocalBisection** satisfies, for an $s \in \{\pm 1\}$,*

$$H_{ij} = sH_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ such that } H_{ij} \neq 0 \text{ or } \frac{|\pi_i - \pi_j|}{n} \geq \rho,$$

where $\rho = (\delta_2 + \omega)/\tilde{\alpha}$.

Lemma D.5 gives guarantees for lines 6-12 of **AgregLocalBisection**. It bounds the error of H by $\rho = (\delta_2 + \omega)/\tilde{\alpha}$. Thus, **AgregLocalBisection** gives an estimate of $H^{(\pi)}$ with accuracy $\rho = (\delta_2 + \omega)/\tilde{\alpha}$. The proof of Lemma D.5 is in appendix F.4.

Remark D.3. The double checking in lines 6-12 of **AgregLocalBisection** is necessary since we may have $j \notin L_i \cup R_i$ but $i \in L_j \cup R_j$. This asymmetrical scenario may occur when one of the sets L_i, R_i is empty, e.g., when $j \in R_i^*$ but $R_i = \emptyset$ we would have $j \notin L_i \cup R_i$. Such empty sets cause a loss of information, but we circumvent this by a simple double checking (first in lines 7-8, second in lines 9-10). The empty sets come from the roughness of **LocalBisection**; see Remark D.2.

D.4 Proof of Proposition C.7

The latent positions π_j/n are denoted by $x_j = \pi_j/n$ for all $j \in [n]$. Given a matrix $H_1 \in \{-1, 0, 1\}^{n \times n}$, let S_{H_1} the set of all undetermined pairs $i < j$ by H_1 , that is:

$$S_{H_1} = \{(i, j) \in [n]^2 : i < j \text{ and } (H_1)_{ij} = 0\}. \quad (48)$$

For any $(i, j) \in S_{H_1}$, **LocalRefine** computes the sets $L_{ij}, R_{ij} \subset [n]$, which are equal to

$$L_{ij} = \{k : (H_1)_{ik} = (H_1)_{jk} = 1\} \quad \text{and} \quad R_{ij} = \{k : (H_1)_{ik} = (H_1)_{jk} = -1\}. \quad (49)$$

For the purpose of the current analysis (and also for the future analysis of **LocalRefineWS**), we encapsulate the subsequent instructions of **LocalRefine** into a sub-algorithm **Test**,

Algorithm 7 Test

Require: $(i, j, B_{ij}^-, B_{ij}^+, A, \sigma)$

Ensure: $H_{ij} \in \{-1, 0, 1\}$

- 1: $l_{ij} = \sum_{k \in B_{ij}^-} A_{ik} - A_{jk}$
 - 2: $r_{ij} = \sum_{k \in B_{ij}^+} A_{ik} - A_{jk}$
 - 3: **if** $|l_{ij}| \geq 5\sigma\sqrt{n \log(n)}$ **then**
 - 4: $H_{ij} = -\text{sign}(l_{ij})$ and $H_{ji} = -H_{ij}$
 - 5: **else**
 - 6: **if** $|r_{ij}| \geq 5\sigma\sqrt{n \log(n)}$ **then**
 - 7: $H_{ij} = \text{sign}(r_{ij})$ and $H_{ji} = -H_{ij}$
 - 8: **end if**
 - 9: **end if**
-

so that, LocalRefine uses Test for the inputs

$$B_{ij}^- = L_{ij} \quad \text{and} \quad B_{ij}^+ = R_{ij} . \quad (50)$$

We will sometimes write $B_{ij}^{-\epsilon}$ and B_{ij}^{ϵ} for an unknown $\epsilon \in \{\pm\}$, with the convention

$$-\epsilon = \begin{cases} - & \text{if } \epsilon = + , \\ + & \text{if } \epsilon = - . \end{cases}$$

To have guarantees for the output H_{ij} of Test, we will use the general Lemma D.6, which states that: if the two input sets B_{ij}^-, B_{ij}^+ satisfy the inclusions (51), and the noise E is a realisation of the event (52), then, the output H_{ij} will have the desired accuracy (53), up to a sign change $\epsilon \in \{\pm\}$ which is the orientation of the sets $B_{ij}^{-\epsilon}, B_{ij}^{\epsilon}$ in (51). To read the conditions (51), we recall some notations: L_{ij}^*, R_{ij}^* are respectively the left and right sets of i, j with respect to the ordering π , that is $L_{ij}^* = \{\ell : x_\ell < x_i \wedge x_j\}$ and $R_{ij}^* = \{\ell : x_\ell > x_i \vee x_j\}$. For any $\lambda > 0$, we also use the left and right λ sub-sets of i, j , which are defined by

$$sL_{ij}^*(\lambda) = \{\ell : x_\ell < (x_i \wedge x_j) - \lambda\} \quad \text{and} \quad sR_{ij}^*(\lambda) = \{\ell : x_\ell > (x_i \vee x_j) + \lambda\} .$$

Lemma D.6. *For any constants $0 < \gamma, r'$ and $0 \leq r''$, the following holds for any (Robinson) matrix F in $\text{SCA}[\gamma, r', r'']$ (as defined in Assumption 4.2), and for any $\tilde{\rho} \in [0, r']$. If i, j satisfy $|x_i - x_j| \leq \tilde{\rho}$, and B_{ij}^-, B_{ij}^+ are such that, for an $\epsilon \in \{\pm\}$,*

$$\begin{aligned} sL_{ij}^*(r'') \subset B_{ij}^{-\epsilon} \subset L_{ij}^* & \quad \text{when} \quad x_i \wedge x_j \geq 1/2 - \tilde{\rho}/2 , \\ sR_{ij}^*(r'') \subset B_{ij}^{\epsilon} \subset R_{ij}^* & \quad \text{when} \quad x_i \vee x_j \leq 1/2 + \tilde{\rho}/2 , \end{aligned} \quad (51)$$

then, conditionally on the event

$$\mathcal{E}_{B_{ij}^\pm} = \left\{ \max_{\epsilon \in \{\pm\}} \frac{1}{\sqrt{2\#B_{ij}^\epsilon}} \left| \sum_{\ell \in B_{ij}^\epsilon} (E_{j\ell} - E_{i\ell}) \right| \leq \sqrt{10 \log(n)} \right\} \quad (52)$$

the output H_{ij} of Test satisfies

$$H_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{whenever} \quad |x_i - x_j| \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}} . \quad (53)$$

The proof of Lemma D.6 is in appendix G.

Let us now check that the assumptions of Lemma D.6 are satisfied by the inputs $B_{ij}^- = L_{ij}$ and $B_{ij}^+ = R_{ij}$. By assumption, the input H_1 of `LocalRefine` is deterministic, or more generally, is independent of A . Therefore, the sets L_{ij}, R_{ij} , which are induced by H_1 in (49), are also independent of A . Then, Lemma D.7 ensures that, with high probability, the events (52) for $B_{ij}^- = L_{ij}$ and $B_{ij}^+ = R_{ij}$ hold, simultaneously, for all $i, j \in S_{H_1}$ (recall that the set S_{H_1} is defined in (48)). The proof of Lemma D.7 is in appendix G.

Lemma D.7. *If the inputs B_{ij}^-, B_{ij}^+ of `Test` are deterministic (or more generally, are independent of A), then, with probability at least $1 - 4/n^3$, the following event holds.*

$$\bigcap_{(i,j) \in S_{H_1}} \mathcal{E}_{B_{ij}^\pm} = \left\{ \max_{(i,j) \in S_{H_1}, \epsilon \in \{\pm\}} \frac{1}{\sqrt{2\#B_{ij}^\epsilon}} \left| \sum_{k \in B_{ij}^\epsilon} (E_{ik} - E_{jk}) \right| \leq \sqrt{10 \log(n)} \right\}$$

Lemma D.8 show that, all pairs L_{ij}, R_{ij} share the same orientation as the one encoded by the input H_1 of `LocalRefine`. Therefore, with respect to a same $\epsilon \in \{\pm\}$, the inclusions (51) are fulfilled by the sets $B_{ij}^- = L_{ij}$ and $B_{ij}^+ = R_{ij}$ for all $i, j \in S_{H_1}$.

Lemma D.8. *Let $r'' \in [0, 1)$, $\epsilon \in \{\pm\}$ and $\tilde{\rho} \in [0, r'']$. Assume that the input H_1 of `LocalRefine` has the following accuracy*

$$(H_1)_{k\ell} = \epsilon H_{k\ell}^{(\pi)} \quad \text{for all } k, \ell \in [n] \text{ such that } (H_1)_{k\ell} \neq 0 \text{ or } |x_k - x_\ell| \geq \tilde{\rho} . \quad (54)$$

Then, writing $B_{ij}^- = L_{ij}$ and $B_{ij}^+ = R_{ij}$, the following inclusions hold for all $(i, j) \in S_{H_1}$,

$$sL_{ij}^*(r'') \subset B_{ij}^{-\epsilon} \subset L_{ij}^* \quad \text{and} \quad sR_{ij}^*(r'') \subset B_{ij}^\epsilon \subset R_{ij}^* .$$

The proof of Lemma D.8 is in appendix G.3. Note that the wanted inclusions (51) are weaker than the conclusion of Lemma D.8 which holds without any restriction on the x_i, x_j . In fact, the weaker requirement (51) will be relevant only later, for analyzing the more complex algorithm `LocalRefineWS`.

The next lemma gives us the proximity condition of Lemma D.6, that is $|x_i - x_j| \leq \tilde{\rho}$ for all $i, j \in S_{H_1}$. The proof is in appendix G.3.

Lemma D.9. *Let $\tilde{\rho} \in (0, 1]$. If a matrix $H_1 \in \{-1, 0, 1\}^{n \times n}$ satisfies, for an $\epsilon \in \{\pm\}$ and all k, ℓ ,*

$$(H_1)_{k\ell} = \epsilon H_{k\ell}^{(\pi)} \quad \text{whenever } (H_1)_{k\ell} \neq 0 \text{ or } |x_k - x_\ell| \geq \tilde{\rho} ,$$

then, we have $|x_i - x_j| < \tilde{\rho}$ for all $(i, j) \in S_{H_1}$, where S_{H_1} is defined in (48).

We have proved that the conditions of Lemma D.6 are, with probability $1 - 4/n^3$, satisfied uniformly for all $(i, j) \in S_{H_1}$. Therefore, the conclusion of Lemma D.6 holds for all $(i, j) \in S_{H_1}$, and with respect to the orientation $\epsilon \in \{\pm\}$ encoded by the input H_1 of `LocalRefine` in (34). The proof of Proposition C.7 is complete.

D.5 Proof of Proposition C.9

We recall that $H_{ij}^{(\pi)} = 1 - 2\mathbb{1}_{\pi_i < \pi_j}$ for all $i, j \in [n]$. Let $H \in \{-1, 0, 1\}^{n \times n}$ be a comparison matrix with an error less than some $\nu > 0$, that is, h satisfies (33) for some $s \in \{\pm\}$. Without the loss of generality, we assume that $s = +$. This gives

$$H_{ij} = H_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ such that } |\pi_i - \pi_j| \geq \nu n. \quad (55)$$

To obtain $L_{\max}(\pi^h, \pi) \leq 2\nu$ for $\pi^H = (\pi_1^H, \dots, \pi_n^H)$ defined in (38), it is sufficient to prove that

$$\pi_j^H > \pi_i^H \quad \text{for all } i, j \in [n] \text{ such that } \pi_j \geq \pi_i + 2\nu n .$$

By definition of π^H , we have to show that

$$S_j^H - S_i^H = \sum_{k=1}^n H_{jk} - H_{ik} \geq 1 \quad \text{for all } i, j \in [n] \text{ such that } \pi_j \geq \pi_i + 2\nu n. \quad (56)$$

We now prove (56). Let a pair $i, j \in [n]$ such that $\pi_j \geq \pi_i + 2\nu n$. We introduce the following partition of the ordering π ,

$$\begin{aligned} I_1 &= [1, \pi_i - \nu n], & I_2 &= (\pi_i - \nu n, \pi_i + \nu n), & I_3 &= [\pi_i + \nu n, \pi_j - \nu n], \\ I_4 &= (\pi_j - \nu n, \pi_j + \nu n), & I_5 &= [\pi_j + \nu n, n], \end{aligned}$$

assuming that $\pi_i > \nu n$ and $\pi_j + \nu n \leq n$ (the other cases, where $\pi_i \leq \nu n$ or $\pi_j + \nu n > n$, can be similarly analyzed with a slight adaptation of the partition). We define the associated partition of indices $R_s = \{k \in [n] : \pi_k \in I_s\}$ for $s \in [5]$.

- For $\pi_k \in I_1 \cup I_5$, we have $|\pi_i - \pi_k| \wedge |\pi_j - \pi_k| \geq \nu n$, hence (55) gives $H_{ik} = H_{ik}^{(\pi)}$ and $H_{jk} = H_{jk}^{(\pi)}$ for all $k \in R_1 \cup R_5$. Since $H_{ik}^{(\pi)} = H_{jk}^{(\pi)}$ for all $k \in R_1 \cup R_5$, we have

$$\sum_{k \in R_1 \cup R_5} H_{jk} - H_{ik} = 0 .$$

- For $\pi_k \in I_2$, we have $\pi_j - \pi_k \geq \nu n$, hence $H_{jk} = H_{jk}^{(\pi)} = 1$ for all $k \in R_2$, and

$$\sum_{k \in R_2} H_{jk} - H_{ik} = 1 + \sum_{k \in R_2, k \neq i} H_{jk} - H_{ik} \geq 1 ,$$

where we used $H_{ii} = 0$ in the equality, and $H_{ik} \in \{-1, 0, 1\}$ in the inequality. A similar reasoning yields $\sum_{k \in R_4} H_{jk} - H_{ik} \geq 1$, as R_2 and R_4 are symmetric.

- For $\pi_k \in I_3$, we have $|\pi_i - \pi_k| \wedge |\pi_j - \pi_k| \geq \nu n$, hence $H_{ik} = H_{ik}^{(\pi)} = -1$ and $H_{jk} = H_{jk}^{(\pi)} = 1$ for all $k \in R_3$. It follows that

$$\sum_{k \in R_3} H_{jk} - H_{ik} \geq 2 \#R_3 ,$$

where $\#R_3$ is the cardinal number of the set R_3 .

Gathering the (bullet points) above, we obtain (56). This concludes the proof.

E Proof of Lemma D.1

Under the assumption that $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, \sigma]$, Lemma D.1 gave an upper bound on the errors $|D_{ij}^* - (\hat{D})_{ij}|$ for close i, j , in terms of $\tilde{\beta}, \sigma, n$. In fact, Lemma D.1 follows from the next result which is valid for any F , in particular for those that do not satisfy $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, \sigma]$. Recall the notation $m_i \in \{1, \dots, n\}$, $m_i \neq i$, which denotes a nearest neighbor of i according to the distance D^* , that is $m_i \in \operatorname{argmin}_{t: t \neq i} D_{it}^*$.

Lemma E.1. *Consider the observation model $A = F_\pi + \sigma E$, where E has independent (up to symmetries) sub-Gaussian entries, with zero means and variance proxies all smaller than 1. Then, for any $4 \leq n$, the estimator \hat{D} described in appendix B.1 satisfies, with probability $1 - 1/n^4$,*

$$\max_{i, j \in [n]} \left| (D_{ij}^*)^2 - (\hat{D})_{ij}^2 \right| \lesssim \frac{|F|_{2, \infty}}{\sqrt{n}} \max_{i \in [n]} D_{im_i}^* + \left[\sigma + \frac{|F|_{2, \infty}}{\sqrt{n}} \right] \sigma \sqrt{\frac{\log(n)}{n}}$$

where the notation $a \lesssim b$ (for any real numbers a, b) means that $a \leq Cb$ for a numerical constant C .

The error bound has two parts: the term $D_{im_i}^*$ is the l_2 -distance between i and its nearest neighbor m_i . It is a bias-type term that comes from the nearest neighbor approximation (25). The second term that contains $\sigma \sqrt{\log(n)/n}$ is an upper bound on the fluctuations of the sub-Gaussian noise.

Proof of Lemma D.1. To bound $|D_{ij}^* - (\hat{D})_{ij}|$, we first use the inequality $|a - b| \leq \sqrt{|a^2 - b^2|}$, then apply Lemma E.1, and finally use the inequality $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$, which are valid for any non-negative real numbers a, b . For $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, \sigma]$, the bias term can be upper bounded by $D_{im_i}^* \leq \tilde{\beta}/n$. For $F \in [0, 1]^{n \times n}$, we can check that $|F|_{2, \infty}/\sqrt{n} \leq 1$. Thus Lemma D.1 follows. \square

Lemma E.1 is an extension of the work in [Issartel, 2021] to real-valued matrices A with sub-Gaussian noise. The proof of Lemma E.1 follows the same steps as in [Issartel, 2021]. For completeness, elements of proof can be found below.

E.1 Elements of proof for Lemma E.1

The next lemma shows that the inner product $\langle A_i, A_j \rangle_n$ (between two different columns) is a consistent estimator of the crossed term $\langle F_{\pi_i}, F_{\pi_j} \rangle_n$ in (22).

Lemma E.2. *Consider the observation model $A = F_\pi + \sigma E$, where E has independent (up to symmetries) sub-Gaussian entries, with zero means and variance proxies all smaller than 1. Then, for any $4 \leq n$, we have, with probability $1 - 1/n^4$,*

$$\forall i, j \in [n], i < j: \quad \left| \langle A_i, A_j \rangle_n - \langle F_{\pi_i}, F_{\pi_j} \rangle_n \right| \leq C \left[\frac{|F|_{2, \infty}}{\sqrt{n}} + \sigma \right] \sigma \sqrt{\frac{\log n}{n}},$$

where C is a numerical constant, and we used the notation $|F|_{2, \infty} = \max_{i \in [n]} \|F_i\|_2$, with $\|F_i\|_2^2 = \sum_{j \in [n]} F_{ij}^2$.

Lemma E.2 states that, with high probability, $\langle A_i, A_j \rangle_n$ is close to its mean $\langle F_{\pi_i}, F_{\pi_j} \rangle_n$. The proof in appendix E.2 is based on standard concentration bounds.

The strategy of the distance estimation is to estimate separately the crossed term and the two quadratic terms in (22). Combining Lemma E.2 (for the crossed term) and the nearest neighbor approximation (25) (for the quadratic terms), it is possible to prove the following high probability bound for \widehat{D} :

$$\max_{i,j \in [n]} \left| (D_{ij}^*)^2 - (\widehat{D})_{ij}^2 \right| \lesssim \frac{|F|_{2,\infty}}{\sqrt{n}} \max_{i \in [n]} D_{im_i}^* + \left[\frac{|F|_{2,\infty}}{\sqrt{n}} + \sigma \right] \sigma \sqrt{\frac{\log(n)}{n}},$$

which is the bound in Lemma E.1.

E.2 Proof of Lemma E.2

Let $i, j \in [n]$, $i < j$. For the model $A = F_\pi + \sigma E$, we have

$$\langle A_i, A_j \rangle_n - \langle F_{\pi_i}, F_{\pi_j} \rangle_n = \sigma \langle F_{\pi_i}, E_j \rangle_n + \sigma \langle E_i, F_{\pi_j} \rangle_n + \sigma^2 \langle E_i, E_j \rangle_n. \quad (57)$$

We control each term separately. The term $\langle F_{\pi_i}, E_j \rangle_n$ is a linear combination of $n - 2$ (centered) independent sub-Gaussian random variables, with variance proxies smaller than 1. Using Hoeffding's inequality, we obtain

$$|\langle F_{\pi_i}, E_j \rangle_n| \leq C_1 |F|_{2,\infty} \frac{\sqrt{\log n}}{n},$$

with probability $1 - 1/n^6$, where C_1 is a numerical constant, and $n \geq 4$. The other term $\langle E_i, F_{\pi_j} \rangle_n$ of (57) is equal to $\langle F_{\pi_j}, E_i \rangle_n$ and admits the same upper bound as above.

For $k \neq i, j$, the random variable $E_{ik}E_{jk}$ is the product of two sub-Gaussian r.v., and hence is a sub-exponential r.v.. The scalar product $\langle E_i, E_j \rangle_n$ is therefore a sum of independent sub-exponential r.v. (divided by n), and Bernstein's inequality yields

$$|\langle E_i, E_j \rangle_n| \leq C_2 \sqrt{\frac{\log n}{n}},$$

with a probability greater than $1 - 1/n^6$, where C_2 is a numerical constant.

Taking a union bound over all $i, j \in [n]$, $i < j$, we obtain the uniform control of Lemma E.2 which holds with probability at least $1 - 1/n^4$.

F Proofs of Lemma D.2 to D.5

We prove Lemma D.2, D.3, D.4, D.5 in section F.1, F.2, F.3, F.4, respectively.

F.1 Proof of Lemma D.2 (LocalBisection with ideal input)

Fix $i \in [n]$. Let \mathcal{G}_i the graph build by `LocalBisection`($i, D, \delta_1, \delta_2, \delta_3$), where $\delta_1, \delta_2, \delta_3$ are and positive real numbers, and $D_{ij} = |x_i - x_j|$ for all i, j , using the notation

$$x_j = \pi_j/n \quad \text{for all } j \in [n]. \quad (58)$$

We will use Lemma F.1, F.2, F.3 to prove Lemma D.2.

Lemma F.1. *If $\delta_1 \leq \delta_2$, all nodes of a connected component of \mathcal{G}_i are on the same side of i .*

Lemma F.2. *If $\delta_1 \geq 1/n$, all k such that $x_k \leq x_i - \delta_2$ (resp. $x_k \geq x_i + \delta_2$) are in a same connected component of \mathcal{G}_i .*

Lemma F.3. *Under the hypothesis of Lemma F.2, and if $\delta_2 \leq \delta_3 \leq 1/4$, there exist at least one and at most two connected components of \mathcal{G}_i including $k \in [n]$ such that $|x_k - x_i| \geq \delta_3$.*

By Lemma F.3, the rule $D_{ik} \geq \delta_3$ in (27) of `LocalBisection` is fulfilled by at least one and at most two connected components of \mathcal{G}_i , hence `LocalBisection` outputs at least one and at most two sets (of $[n]$). We denote these outputs by $G_i^{(1)}$ and $G_i^{(2)}$, using the convention $G_i^{(1)} \neq \emptyset$, while $G_i^{(2)}$ may be empty. This gives the point 1 of Lemma D.2 follows.

Assume that $G_i^{(2)} = \emptyset$. By Lemma F.1, $G_i^{(1)}$ is on one side of i , and thus $G_i^{(1)} \subset L_i^*$ or $G_i^{(1)} \subset R_i^*$. By symmetry, we only focus on the case $G_i^{(1)} \subset R_i^*$. Then, Lemma F.3 ensures that, $G_i^{(1)}$ includes $k \in R_i^*$ at distance (at least) δ_3 , i.e., there exists $k \in G_i^{(1)}$ such that $x_k - x_i \geq \delta_3$. Since $\delta_3 \geq \delta_2$, we obtain $x_k - x_i \geq \delta_2$, hence $k \in sR_i^*(\delta_2)$. Combining with Lemma F.2 we conclude that

$$\emptyset \neq sR_i^*(\delta_2) \subset G_i^{(1)} .$$

The point 2 of Lemma D.2 is proved.

If $G_i^{(2)} \neq \emptyset$, we can similarly prove the point 3. The proof of Lemma D.2 is complete. \square

Lemma F.1, F.2, F.3 are proved below.

Proof of Lemma F.1. If two nodes are connected (by one edge), then they are on the same side of i . Indeed, let $k, \ell \in [n]$, $k, \ell \neq i$, be a pair of connected nodes, then (by construction of \mathcal{G}_i) the pair k, ℓ satisfies the rule (26), i.e., we have $|x_k - x_\ell| \leq \delta_1$ and $|x_i - x_\ell| \vee |x_i - x_k| \geq \delta_2$. If $\delta_2 > \delta_1$, the objects k, ℓ are necessarily on the same side of i . If $\delta_2 = \delta_1$, the same conclusion holds since $\min_{s \in [n], s \neq i} |x_i - x_s| = 1/n > 0$.

We readily deduce from the above that all nodes of a connected component are on a same side. Let $k, k' \in [n]$ be in a same connected component, then there exists a path (of connected nodes) going from k to k' . Since any two consecutive nodes along the path are on the same side, the extremities k and k' are necessarily on a same side too. \square

Proof of Lemma F.2. Let k such that $x_k \leq x_i - \delta_2$ (the symmetric case $x_k \geq x_i + \delta_2$ is omitted). If there do not exist $k' \neq k$ such that $x_{k'} \leq x_i - \delta_2$, then the statement of the lemma holds (trivially). In the following, assume that such a k' exists. Here we assume (for simplicity) that $\pi = id$, that is, the notation (58) becomes $x_j = j/n$ for all j . We only focus on the case $k' < k$ (the other case $k < k'$ being symmetric). The objects $k-1, k$ are consecutive objects in the ordering π , and so $|x_{k-1} - x_k| = 1/n \leq \delta_1$. Since they are also at distance δ_2 from i , the rule (26) holds for k and $\ell = k-1$. This means that k and $k-1$ are connected by an edge in the graph. By induction, every pair of consecutive nodes along the path $k', k'+1, \dots, k-1, k$ is connected by an edge. Thus, k, k' belong to a same connected component. \square

Proof of Lemma F.3. By Lemma F.2, all objects on one side of i , which are at distance δ_2 from i , belong to a same connected component. The graph \mathcal{G}_i has therefore at most two connected components that contain a k such that $|x_i - x_k| \geq \delta_3$ for $\delta_3 \geq \delta_2$.

Let us show the existence. Here we assume (for simplicity) that $\pi = id$, so that the notation (58) becomes $x_j = j/n$ for all j . Observe that $|x_i - x_1| \geq 1/4$ or $|x_n - x_i| \geq 1/4$. Therefore, there exists $k \in \{1, n\}$ such that $|x_i - x_k| \geq \delta_3$ for $\delta_3 \leq 1/4$. \square

F.2 Proof of Lemma D.3 (LocalBisection)

At a high level, the proof follows the same steps as for Lemma D.2 (appendix F.1). Fix $i \in [n]$, and let \mathcal{G}_i the graph build by `LocalBisection`($i, D, \delta_1, \delta_2, \delta_3$), where D is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$. Lemma D.3 follows from the three next lemmas, which are generalizations of the three lemmas used in the proof of Lemma D.2. We use again the notation $x_j = \pi_j/n$ from (58).

Lemma F.4. *All nodes of a connected component of \mathcal{G}_i are on the same side of i , if*

$$\delta_1 \geq \omega + \frac{\tilde{\beta}}{n}, \quad \delta_2 > \omega + \frac{\tilde{\beta}}{\tilde{\alpha}}(\delta_1 + \omega), \quad r \geq \frac{n^{-1} \vee (\delta_1 + \omega) \vee (\delta_2 + \omega)}{1 \wedge \tilde{\alpha}}.$$

Lemma F.5. *Under the hypotheses of Lemma F.4, all ℓ such that $x_\ell \leq x_i - \rho$ (respectively $x_\ell \geq x_i + \rho$) are in a same connected component of \mathcal{G}_i .*

Lemma F.6. *Under the hypothesis of Lemma F.5, and if $n \geq 4$ and $\omega + \tilde{\beta}\rho < \delta_3 \leq (r \wedge (\tilde{\alpha}/4)) - \omega$, there exist at least one and at most two connected components of \mathcal{G}_i including $k \in [n]$ such that $D_{ki} \geq \delta_3$.*

Using Lemma F.4, F.5, F.6, we obtain Lemma D.3 in the same way as we did for Lemma D.2 (in appendix F.1). \square

Lemma F.4, F.5, F.6 are proved below.

F.2.1 Proof of Lemma F.4, F.5, F.6

Using the assumption that D is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$, we first derive some relations between the inputs D_{ij} and the latent distances $|x_i - x_j|$. Lemma F.7 gives the useful relations (59-62).

Lemma F.7. *Under the hypotheses of Lemma F.4, the following conditions are fulfilled*

$$D_{kl} \leq \delta_1 \implies |x_k - x_\ell| \leq \kappa \tag{59}$$

$$|x_i - x_\ell| \leq \kappa \implies D_{i\ell} < \delta_2 \tag{60}$$

$$D_{i\ell} < \delta_2 \implies |x_i - x_\ell| < \rho \tag{61}$$

$$D_{\ell\ell_c} \leq \delta_1 \tag{62}$$

for all i, k, ℓ , where ℓ_c is the consecutive object after ℓ in the ordering π (i.e., $\pi_{\ell_c} = \pi_\ell + 1$), and

$$\kappa = \frac{\delta_1 + \omega}{\tilde{\alpha}} \quad \text{and} \quad \rho = \frac{\delta_2 + \omega}{\tilde{\alpha}}.$$

Proof of Lemma F.7. If $D_{k\ell} \leq \delta_1$, where D is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ and $\delta_1 \leq r$, then the \mathcal{LDE} (local distance equivalence) in Assumption (4.1) gives

$$|x_k - x_\ell| \leq \frac{D_{k\ell} + \omega}{\tilde{\alpha}} \leq \frac{\delta_1 + \omega}{\tilde{\alpha}} = \kappa .$$

Similarly, if $|x_i - x_\ell| \leq \kappa$, with $\kappa \leq r$, then $D_{i\ell} \leq \tilde{\beta}\kappa + \omega < \delta_2$ by the \mathcal{LDE} .

If $D_{i\ell} < \delta_2$, with $\delta_2 \leq r$, then $|x_i - x_\ell| < (\delta_2 + \omega)/\tilde{\alpha} = \rho$.

We have $|x_\ell - x_{\ell_c}| = 1/n$, with $1/n \leq r$, so $D_{\ell\ell_c} \leq (\tilde{\beta}/n) + \omega \leq \delta_1$.

Lemma F.7 follows. \square

We are now ready to prove Lemma F.4, F.5, F.6.

Proof of Lemma F.4. Assume that the rule (26) holds, i.e., $D_{k\ell} \leq \delta_1$ and $D_{ik} \vee D_{i\ell} \geq \delta_2$. Then, (59-60 yield $|x_k - x_\ell| \leq \kappa$ and $|x_i - x_\ell| \vee |x_i - x_k| > \kappa$. Therefore, k and ℓ are necessarily on the same side of i . As in the proof of Lemma F.1, we can deduce that all nodes of a connected component of \mathcal{G}_i are on the same side of i . \square

Proof of Lemma F.5. Let ℓ, ℓ' such that $x_\ell \leq x_i - \rho$ and $x_{\ell'} \leq x_i - \rho$. The relation (61) yields $D_{i\ell} \wedge D_{i\ell'} \geq \delta_2$. We only focus on the case $\pi_\ell < \pi_{\ell'}$. For the object ℓ_c that is consecutive after ℓ in the ordering π , the relation (62) gives $D_{\ell\ell_c} \leq \delta_1$. Therefore, the rule (26) holds for ℓ and $k = \ell_c$. This means that the nodes ℓ, ℓ_c are connected by an edge. As in Lemma F.2, we can conclude by induction. Thus, ℓ and ℓ' are in a same connected component of \mathcal{G}_i . \square

Proof of Lemma F.6. By Lemma F.5, all objects k on one side of i , which are at distance $|x_k - x_i| \geq \rho$, belong to a same connected component. The graph \mathcal{G}_i has therefore at most two connected components that contain a k such that $|x_k - x_i| \geq \rho$. As proved below, $D_{ki} \geq \delta_3$ implies $|x_k - x_i| \geq \rho$, so the graph \mathcal{G}_i has at most two connected components including a $k \in [n]$ such that $D_{ki} \geq \delta_3$.

By contraposition: if $|x_i - x_k| < \rho$ (with $\rho \leq r$), then the \mathcal{LDE} Assumption (4.1) gives $D_{ki} < \tilde{\beta}\rho + \omega \leq \delta_3$ for $\tilde{\beta}\rho + \omega \leq \delta_3$.

Let us show the existence. Assume (for simplicity) that $\pi = id$, so that $x_j = j/n$ for all j . If $D_{ki} < \delta_3$ (with $\delta_3 \leq r$), the \mathcal{LDE} yields $|x_k - x_i| < (\delta_3 + \omega)/\tilde{\alpha} \leq 1/4$ for $\delta_3 \leq (\tilde{\alpha}/4) - \omega$. But we know that $|x_i - x_1| \geq 1/4$ or $|x_i - x_n| \geq 1/4$, so $D_{ki} \geq \delta_3$ for some $k \in \{1, n\}$. \square

F.3 Proof of Lemma D.4 (Orientation)

We assume (for simplicity) that $\pi = id$, so that the notation (58) becomes $x_j = j/n$ for all j .

Line 5 of Orientation: This step takes an index i^* in $V_{\neq\emptyset}$, so we have to show that $V_{\neq\emptyset} \neq \emptyset$. We do so by proving that $i_o = \lfloor n/2 \rfloor$ belongs to $V_{\neq\emptyset}$. By definition, $i_o \in V_{\neq\emptyset}$ means that $G_{i_o}^{(1)} \neq \emptyset$ and $G_{i_o}^{(2)} \neq \emptyset$, i.e. **LocalBisection** outputs two (non-empty) connected components (of the graph \mathcal{G}_{i_o}). This scenario happens only when the rule (27) is met by two objects k

and k' (i.e. $D_{ki} \geq \delta_3$ and $D_{k'i} \geq \delta_3$) which come from different components of \mathcal{G}_{i_o} . Therefore, it is enough to show that the rule (27) is satisfied for $k = 1$ and $k' = n$, i.e.:

$$D_{i_o1} \wedge D_{i_on} \geq \delta_3 . \quad (63)$$

Indeed, the objects 1 and n are on different sides of i_o (in the ordering $\pi = id$), and thus (by Lemma F.4) they do not belong to a same connected component.

Proof of (63) by contradiction: If $D_{i_o1} < \delta_3$ (with $\delta_3 \leq r$), the \mathcal{LDE} Assumption 4.1 would yield $x_{i_o} - x_1 < (\delta_3 + \omega)/\tilde{\alpha} \leq 1/8$. But this contradicts the fact that $x_{i_o} - x_1 = i_o/n - 1/n \geq (1/2 - 1/n) - 1/n \geq 1/4$ (for $n \geq 8$). The same applies to $D_{i_on} < \delta_3$. Thus (63) follows. \square

Lines 12 and 14: We recall that the inputs $G_i^{(1)}, G_i^{(2)}$ are equal to the outputs of `LocalBisection`.

Let $i^*, i \in V_{\neq \emptyset}$, $i \neq i^*$. Since $i^* \in V_{\neq \emptyset}$, both sets $L_{i^*} = G_{i^*}^{(1)}$ and $R_{i^*} = G_{i^*}^{(2)}$ are non-empty, and we can apply the point 3 of Lemma D.3. Combining with the orientation (45) we obtain

$$\emptyset \neq sL_{i^*}^*(\rho) \subset L_{i^*} \subset L_{i^*}^* \quad \text{and} \quad \emptyset \neq sR_{i^*}^*(\rho) \subset R_{i^*} \subset R_{i^*}^* . \quad (64)$$

Thus i^* fulfills the wanted property (46) of the lemma.

For $i \in V_{\neq \emptyset}$, both sets $G_i^{(1)}$ and $G_i^{(2)}$ are non-empty, and again we apply the point 3 of Lemma D.3. This gives

$$\emptyset \neq sL_i^*(\rho) \subset G_i^{(u)} \subset L_i^* \quad \text{and} \quad \emptyset \neq sR_i^*(\rho) \subset G_i^{(v)} \subset R_i^* , \quad (65)$$

for some $u, v \in [2]$, $u \neq v$. We now check if the algorithm successfully rearranges this pair of sets $G_i^{(1)}, G_i^{(2)}$, so as to align them the i^* -direction in (64). There are two scenarios: either $G_i^{(1)}, G_i^{(2)}$ are (respectively) on the left and right sides of i , or vice-versa. We can focus on one case, say $G_i^{(1)}$ and $G_i^{(2)}$ are on the left and the right sides of i , respectively. This orientation and (65) yield

$$\emptyset \neq sL_i^*(\rho) \subset G_i^{(1)} \subset L_i^* \quad \text{and} \quad \emptyset \neq sR_i^*(\rho) \subset G_i^{(2)} \subset R_i^* . \quad (66)$$

The algorithm tests (at most) four intersections, between $G_i^{(u)}$, $u \in [2]$, and the reference sets L_{i^*} and R_{i^*} , and the algorithm stops as soon as an intersection is found empty. We will see that the order of testing does not matter. If it finds the null intersections $G_i^{(1)} \cap R_{i^*} = \emptyset$ or $G_i^{(2)} \cap L_{i^*} = \emptyset$, in both cases it sets $L_i = G_i^{(1)}$ and $R_i = G_i^{(2)}$. Then, (66) will imply

$$\emptyset \neq sL_i^*(\rho) \subset L_i \subset L_i^* \quad \text{and} \quad \emptyset \neq sR_i^*(\rho) \subset R_i \subset R_i^* ,$$

and i will satisfy the property (46) of the lemma.

In light of the above, we only need to check two facts: (1) at least one of the two intersections $G_i^{(1)} \cap R_{i^*}$ or $G_i^{(2)} \cap L_{i^*}$ is null, and (2) the two other intersections, which the algorithm (possibly) tests, are not null. For (2), it suffices to observe that the inclusions (64-66) ensure that $1 \in G_i^{(1)} \cap L_{i^*}$ and $n \in G_i^{(2)} \cap R_{i^*}$. Thus, none of these intersections is null. For (1), the relations (64-66) yield $G_i^{(2)} \cap L_{i^*} = \emptyset$ if $i^* < i$, and $G_i^{(1)} \cap R_{i^*} = \emptyset$ otherwise (when $i < i^*$). \square

Lines 23 and 25: Let $i \notin V_{\neq \emptyset}$. The point 1 of Lemma D.3 gives $G_i^{(1)} \neq \emptyset$. Then, by definition of $V_{\neq \emptyset}$, we have $G_i^{(2)} = \emptyset$. Thus $i \in V'_{\neq \emptyset}$. The point 2 of Lemma D.3 ensures that

$G_i^{(1)}$ is on one side of i . Without the loss of generality, we focus here on the case where $G_i^{(1)}$ is on the right side of i . Then, the point 2 of Lemma D.3 yields

$$\emptyset \neq sR_i^*(\rho) \subset G_i^{(1)} \subset R_i^* . \quad (67)$$

For any $k \in V_{\neq \emptyset}$, the associated sets L_k, R_k have already been defined in lines 12 and 14 of **Orientation**. We have shown above that these sets satisfy the property (46), which we recall:

$$\emptyset \neq sL_k^*(\rho) \subset L_k \subset L_k^* \quad \text{and} \quad \emptyset \neq sR_k^*(\rho) \subset R_k \subset R_k^* . \quad (68)$$

We momentarily assume that one of the tests in lines 23 and 25 is satisfied, i.e., there exists $k \in V_{\neq \emptyset}$ such that

$$i \in L_k \cup R_k \quad \text{and} \quad G_i^{(1)} \cap L_k \neq \emptyset \quad \text{and} \quad G_i^{(1)} \cap R_k \neq \emptyset . \quad (69)$$

Let us determine whether $i \in L_k$ or $i \in R_k$. If $k < i$ were true, then (67-68) would yield $G_i^{(1)} \cap L_k = \emptyset$, hence a contradiction with (69). Therefore we have $i < k$. Reading (69) with this new information and with (68), we see that necessarily $i \in L_k$. Thus, Line 23 of the algorithm sets $L_i = \emptyset$ and $R_i = G_i^{(1)}$. Combining with the inclusions (67), we get

$$\emptyset \neq sR_i^*(\rho) \subset R_i \subset R_i^* .$$

Thus, i fulfills the property (46) of the lemma. \square

Proof of (69). Let us show that $k = k_o = \lfloor 3n/4 \rfloor$ fulfills the three parts of (69). We assume (for now) that

$$(1.\text{bis}) \quad x_i < 1/4 , \quad (2.\text{bis}) \quad k_o \in V_{\neq \emptyset} . \quad (70)$$

◦ Using (1.bis), with $n \geq 8$ and $\rho \leq 1/8$, we obtain

$$x_{k_o} - x_i > (3/4 - 1/n) - 1/4 \geq 1/4 \geq \rho ,$$

hence $i \in sL_{k_o}^*(\rho)$. Since (68) holds for any $k \in V_{\neq \emptyset}$, it holds in particular for $k = k_o$ by (2.bis), and we have

$$sL_{k_o}^*(\rho) \subset L_{k_o} . \quad (71)$$

Therefore, $i \in L_{k_o}$. This gives the first part of (69) for $k = k_o$.

◦ The assumption (2.bis) ensures that (67-68) hold for $k = k_o$. It follows from (67-68) that $n \in G_i^{(1)} \cap R_{k_o}$. The third part of (69) is checked.

◦ By definition of $k_o = \lfloor 3n/4 \rfloor$ we have $x_{k_o} = k_o/n \in [3/4 - 1/n, 3/4] \subset [5/8, 3/4]$ for $n \geq 8$. Then, (1.bis) yields

$$x_{\lfloor n/2 \rfloor} - x_i > (1/2 - 1/n) - 1/4 \geq 1/8 \geq \rho ,$$

for $\rho \leq 1/8$. Thus $\lfloor n/2 \rfloor \in sR_i^*(\rho) \subset G_i^{(1)}$, where the inclusion follows directly from (67). Similarly, we have

$$x_{k_o} - x_{\lfloor n/2 \rfloor} \geq 5/8 - 1/2 \geq 1/8 \geq \rho ,$$

hence $\lfloor n/2 \rfloor \in sL_{k_o}^*(\rho) \subset L_{k_o}$, where we used (71). We have proved $\lfloor n/2 \rfloor \in G_i^{(1)} \cap L_{k_o}$, which is the second part of (69) for $k = k_o$. Thus, (69) is proved. \square

Proof of (70). ◦ Let us start with (1.bis) and recall that $i \notin V_{\neq \emptyset}$. Following the same reasoning as above (63), we proceed by contradiction. If we had $D_{i1} \geq \delta_3$, we would have $1 \in G_i^{(1)} \cup G_i^{(2)}$. But, $G_i^{(2)} = \emptyset$ since $i \notin V_{\neq \emptyset}$, hence $1 \in G_i^{(1)}$. This yields the contradiction with (67). We have proved that $D_{i1} < \delta_3$.

Then, the \mathcal{LDE} Assumption 4.1 yields

$$\tilde{\alpha}|x_i - x_1| \leq D_{i1} + \omega < \delta_3 + \omega \leq \tilde{\alpha}/8 ,$$

for $\delta_3 + \omega \leq \tilde{\alpha}/8$, which leads to

$$x_i < x_1 + 1/8 = 1/n + 1/8 \leq 1/4 , \quad (72)$$

for $n \geq 8$. The proof of (1.bis) is complete.

◦ Let us prove (2.bis), that is, **LocalBisection** releases two non empty connected components $G_{k_o}^{(1)}, G_{k_o}^{(2)}$ of the graph \mathcal{G}_{k_o} . As seen above (63), it suffices to show that $D_{k_o1} \wedge D_{k_on} \geq \delta_3$.

By contradiction, if we had $D_{k_o1} < \delta_3$, then, following the same lines as for (72) we would obtain $x_{k_o} < 1/4$. Similarly, if $D_{k_on} < \delta_3$, then $x_{k_o} > 3/4$. Both inequalities $x_{k_o} < 1/4$ and $x_{k_o} > 3/4$ contradict the definition of $k_o = \lfloor 3n/4 \rfloor$ since $x_{k_o} = k_o/n \in [3/4 - 1/n, 3/4] \subset [5/8, 3/4]$ for $n \geq 8$. Therefore, $D_{k_o1} \wedge D_{k_on} \geq \delta_3$.

Thus, (2.bis) is proved. The proof of (70) is complete. \square

Lemma D.4 is proved.

F.4 Proof of Lemma D.5

Without the loss of generality, we assume that $\pi = id$, so that the notation (58) becomes $x_j = \pi_j/n = j/n$ for all j .

◦ **Let $i, j \in [n]$ such that $H_{ij} \neq 0$.** **AgregLocalBisection** defines H_{ij} as follows: it sets $H_{ij} = -1$ if $j \in R_i$ or $i \in L_i$; it sets $H_{ij} = 1$ otherwise, when $j \in L_i$ or $i \in R_i$. Lemma D.4 ensures that, for any $k \in [n]$, if the set L_k (resp. R_k) is non-empty, then it is on the left (resp. right) side of k . It follows that $H_{ij} = H_{ij}^{(\pi)}$, thus **AgregLocalBisection** recovers the comparison $H_{ij}^{(\pi)}$. \square

◦ **Let $i, j \in [n]$ such that $|x_i - x_j| \geq \rho$.** We focus on the case $i < j$ (the symmetric case can be handled similarly). One can see that $j \in sR_i^*(\rho)$. Besides, if $R_i \neq \emptyset$, Lemma D.4 gives $sR_i^*(\rho) \subset R_i$. Therefore, $j \in R_i$ whenever $R_i \neq \emptyset$. **AgregLocalBisection** sets accordingly $H_{ij} = -1$, which matches the true value $H_{ij}^{(\pi)} = -1$.

We have yet to analyze the situation $R_i = \emptyset$. Lemma D.4 ensures that $L_i \cup R_i \neq \emptyset$, so we necessarily have $L_i \neq \emptyset$. Then, Lemma D.4 ensures that L_i is on the left side of i . Since $i < j$, we have $j \notin L_i \cup R_i$. This means that **AgregLocalBisection** will check whether $i \in L_j \cup R_j$ holds or not, and then will set the correct value $H_{ij} = -1 = H_{ij}^{(\pi)}$ iff $i \in L_j$. To prove that $i \in L_j$, we only have to verify that $L_j \neq \emptyset$. (Indeed, if $L_j \neq \emptyset$, Lemma D.4 ensures that $sL_j^*(\rho) \subset L_j$. Besides, we know that $x_i \leq x_j - \rho$, so $i \in sL_j^*(\rho)$. Therefore, $i \in L_j$.) The

verification of $L_j \neq \emptyset$ is similar to previous arguments, and is delayed at the end of section. Thus, `AgregLocalBisection` sets the correct value $H_{ij} = -1$. \square

Verification of $L_j \neq \emptyset$, in the situation $R_i = \emptyset$. This amounts to checking that the `LocalBisection` releases a connected component (of \mathcal{G}_j) which is on the left side of j . As seen before, it is enough to check that the rule (27) is met by j and 1, that is, $D_{j1} \geq \delta_3$.

By contradiction, assume that $D_{j1} < \delta_3$. The \mathcal{LDE} Assumption (4.1) yields $|x_j - x_1| < (\delta_3 + \omega)/\tilde{\alpha} \leq 1/8$, hence $x_j < 1/8 + x_1 = 1/8 + (1/n) \leq 1/4$ for $n \geq 8$. Thus $x_i < 1/4$ (since $x_i < x_j$).

On the other hand, when $R_i = \emptyset$, we have $D_{in} < \delta_3$ (otherwise, `LocalBisection` would release a component of \mathcal{G}_i that is on the right side of i , i.e., $R_i \neq \emptyset$). Then, the \mathcal{LDE} assumption gives $|x_i - x_n| < (\delta_3 + \omega)/\tilde{\alpha} \leq 1/8$, hence $x_i > x_n - 1/8 \geq (1 - 1/n) - 1/8 \geq 3/4$. Thus $x_i > 3/4$, which brings the contradiction with the conclusion of the previous paragraph. \square

G Proof of Lemma D.6 to D.9

G.1 Proof of Lemma D.6

Recall the notation $x_i = \pi_i/n$ for all i . Without the loss of generality, we focus on the case where $x_i < x_j$, and $\epsilon = +$ in (51) (the other cases are symmetric and can be analyzed similarly). Thus, we have $H_{ij}^{(\pi)} = -1$ and we want to prove that $H_{ij} = -1$, under the assumption (51) which becomes here

$$\begin{aligned} sL_{ij}^*(r'') &\subset B_{ij}^- \subset L_{ij}^* && \text{when } x_i \wedge x_j \geq 1/2 - \tilde{\rho}/2, \\ sR_{ij}^*(r'') &\subset B_{ij}^+ \subset R_{ij}^* && \text{when } x_i \vee x_j \leq 1/2 + \tilde{\rho}/2. \end{aligned} \quad (73)$$

To have $H_{ij} = -1$, we only need to check that one of the two following inequalities holds

$$l_{ij} := \sum_{k \in B_{ij}^-} A_{ik} - A_{jk} \geq 5\sigma\sqrt{n \log(n)} \quad \text{or} \quad r_{ij} := \sum_{k \in B_{ij}^+} A_{ik} - A_{jk} \leq -5\sigma\sqrt{n \log(n)}, \quad (74)$$

while none of the two other results $l_{ij} \leq -5\sigma\sqrt{n \log(n)}$ or $r_{ij} \geq 5\sigma\sqrt{n \log(n)}$ is possible.

• *Proof of (74).* We drop the dependency in i, j for convenience and write B^ϵ for $\epsilon \in \{\pm\}$. Then, to prove (74), we need to show that $\sum_{\ell \in B^\epsilon} \epsilon(A_{j\ell} - A_{i\ell}) \geq 5\sigma\sqrt{n \log(n)}$ for an $\epsilon \in \{\pm\}$. We have

$$\begin{aligned} \sum_{\ell \in B^\epsilon} \epsilon(A_{j\ell} - A_{i\ell}) &= \sum_{\ell \in B^\epsilon} \epsilon(F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell}) + \sum_{\ell \in B^\epsilon} \epsilon\sigma(E_{j\ell} - E_{i\ell}) \\ &\geq \sum_{\ell \in B^\epsilon} \epsilon(F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell}) - \sigma \max_{\epsilon \in \{\pm\}} \left| \sum_{\ell \in B^\epsilon} (E_{j\ell} - E_{i\ell}) \right| \\ &> \sum_{\ell \in B^\epsilon} \epsilon(F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell}) - 5\sigma\sqrt{n \log(n)}, \end{aligned} \quad (75)$$

where the last line holds conditionally on the event \mathcal{E}_{B^\pm} of (52), with $\max_{\epsilon \in \{\pm\}} \#B^\epsilon \leq n$ and $\sqrt{2}\sqrt{10} < 5$. Using the inclusions $B^- \subset L_{ij}^*$ and $B^+ \subset R_{ij}^*$ of (73), and the Robinson shape

(1) of F , we get

$$\begin{aligned} F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell} &> 0 \quad \text{for all } \ell \in B^+ , \\ F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell} &< 0 \quad \text{for all } \ell \in B^- . \end{aligned} \quad (76)$$

since $x_i < x_j$ where $x_i = \pi_i/n$ and $x_j = \pi_j/n$. Therefore, for any $\epsilon \in \{\pm\}$, the sum $\sum_{\ell \in B^\epsilon} \epsilon(F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell})$ is composed of non-negative terms, and hence we can lower bound it by any sub-sum. In particular, if we have the following inclusions

$$sL_{ij}^*(r'') := \{\ell : x_\ell < x_i - r''\} \subset B^- \quad \text{and} \quad sR_{ij}^*(r'') := \{\ell : x_\ell > x_j + r''\} \subset B^+ \quad (77)$$

then we can lower bound the last sum in (75) by

$$\begin{aligned} \text{when } \epsilon = -, \quad \sum_{\ell \in B^-} -(F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell}) &\geq \sum_{\ell: x_\ell < x_i - r''} F_{\pi_i \pi_\ell} - F_{\pi_j \pi_\ell} , \\ \text{when } \epsilon = +, \quad \sum_{\ell \in B^+} (F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell}) &\geq \sum_{\ell: x_\ell > x_j + r''} F_{\pi_j \pi_\ell} - F_{\pi_i \pi_\ell} . \end{aligned} \quad (78)$$

To see why (at least) one of the the inclusions (77) holds, one can readily check that $x_i \geq 1/2 - \tilde{\rho}/2$ or $x_j \leq 1/2 + \tilde{\rho}/2$ (since $|x_i - x_j| \leq \tilde{\rho}$ and $x_i < x_j$ by assumption) and then use assumption (73).

To lower bound (78), we use Assumption 4.2. Recall that this assumption only holds when $x_i \geq 1/2 - r'/2$ or $x_j \leq 1/2 + r'/2$, but one of these two inequalities is actually satisfied here (indeed, we have seen above that $x_i \geq 1/2 - \tilde{\rho}/2$ or $x_j \leq 1/2 + \tilde{\rho}/2$, and we have $\tilde{\rho} \leq r'$ by assumption). Therefore, focusing on one of these two symmetric cases, say $x_i \geq 1/2 - r'/2$, Assumption 4.2 yields

$$\sum_{\ell: x_\ell < x_i - r''} F_{\pi_i \pi_\ell} - F_{\pi_j \pi_\ell} \geq \gamma |x_i - x_j| n .$$

Substituting into (78) and then into (75), we obtain

$$\max_{\epsilon \in \{\pm\}} \sum_{\ell \in B^\epsilon} \epsilon(A_{j\ell} - A_{i\ell}) > \gamma |x_i - x_j| n - 5\sigma \sqrt{n \log(n)} .$$

Thus, the test (74) will be satisfied when $\gamma |x_i - x_j| n - 5\sigma \sqrt{n \log(n)}$ is greater than $5\sigma \sqrt{n \log(n)}$. This gives the following condition for a successful test:

$$|x_i - x_j| \geq 10 \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}} .$$

We retrieve the accuracy (53) of Lemma D.6.

• *Proof that neither $l_{ij} \leq -5\sigma \sqrt{n \log(n)}$ or $r_{ij} \geq 5\sigma \sqrt{n \log(n)}$ is possible.* Following the lines in (75), we obtain

$$\sum_{\ell \in B^\epsilon} \epsilon(A_{i\ell} - A_{j\ell}) < \sum_{\ell \in B^\epsilon} \epsilon(F_{\pi_i \pi_\ell} - F_{\pi_j \pi_\ell}) + 5\sigma \sqrt{n \log(n)} , \quad (79)$$

conditionally on the event \mathcal{E}_{B^\pm} . By (76), the last sum in (79) is non-positive, hence

$$\sum_{\ell \in B^\epsilon} \epsilon(A_{i\ell} - A_{j\ell}) < 5\sigma \sqrt{n \log(n)} . \quad (80)$$

Thus, $r_{ij} < 5\sigma \sqrt{n \log(n)}$ and $l_{ij} > -5\sigma \sqrt{n \log(n)}$.

G.2 Proof of Lemma D.7.

By assumption, the sets B_{ij}^- and B_{ij}^+ are independent of A . Therefore, conditionally to B_{ij}^- and B_{ij}^+ , the $2(n-1)$ real numbers $\{E_{i\ell} : \ell \in [n], \ell \neq i\}$ and $\{E_{j\ell} : \ell \in [n], \ell \neq j\}$ are (still) independent sub-Gaussian random variables, with zero means and variance proxies all smaller than 1. Applying a standard concentration inequality [Rigollet and Hütter, 2023, Corollary 1.7], we obtain

$$\mathbb{P} \left[\frac{1}{\sqrt{2\#B_{ij}^\epsilon}} \left| \sum_{\ell \in B_{ij}^\epsilon} (E_{i\ell} - E_{j\ell}) \right| \geq t \mid B_{ij}^\epsilon \right] \leq 2e^{-t^2/2} ,$$

for all $t > 0$ (using the convention $0/0 = 0$ and $\sum_{k \in \emptyset} = 0$). Since this inequality holds for any $B_{ij}^\epsilon \subset [n]$, we have the same upper bound without conditioning (to B_{ij}^ϵ). Taking $t = \sqrt{10 \log(n)}$, and a union bound over all $\epsilon \in \{\pm\}$ and all $(i, j) \in S_{H_1}$, we get $2e^{-t^2/2} \leq 4/n^3$ (using the bound $\#S_{H_1} \leq n^2$ on the cardinal number of S_{H_1}).

G.3 Proof of Lemma D.8 and D.9.

Proof of Lemma D.8. Without the loss of generality, we assume that the assumption on H_1 holds for $\epsilon = +$, so we have $(H_1)_{k\ell} = H_{k\ell}^{(\pi)}$ for all k, ℓ such that $(H_1)_{k\ell} \neq \emptyset$, or $|x_\ell - x_k| \geq \tilde{\rho}$. In particular, for $(i, j) \in S_{H_1}$ we have $(H_1)_{ik} = H_{ik}^{(\pi)}$ whenever $(H_1)_{ik} \neq \emptyset$, and similarly $(H_1)_{jk} = H_{jk}^{(\pi)}$ whenever $(H_1)_{jk} \neq \emptyset$. Combining with the definition (49) of L_{ij}, R_{ij} , we obtain

$$L_{ij} \subset L_{ij}^* \quad \text{and} \quad R_{ij} \subset R_{ij}^* .$$

Since $(H_1)_{ik} = H_{ik}^{(\pi)}$ when $|x_k - x_i| \geq \tilde{\rho}$, and similarly $(H_1)_{jk} = H_{jk}^{(\pi)}$ when $|x_k - x_j| \geq \tilde{\rho}$, we also have for $\tilde{\rho} \leq r''$,

$$sL_{ij}^*(r'') \subset sL_{ij}^*(\tilde{\rho}) \subset L_{ij} \quad \text{and} \quad sR_{ij}^*(r'') \subset sR_{ij}^*(\tilde{\rho}) \subset R_{ij} .$$

Thus, for $B_{ij}^- = L_{ij}$ and $B_{ij}^+ = R_{ij}$, the conclusion of Lemma D.8 follows. (Note that some sets in these inclusions might be empty, e.g., when x_i, x_j are close to an extremity 0 or 1). \square

Proof of Lemma D.9. Let $(i, j) \in S_{H_1}$. We have $(H_1)_{ij} = 0$ by definition of S_{H_1} . Since $(H_1)_{k\ell} = H_{k\ell}^{(\pi)} \neq 0$ when $|x_k - x_\ell| \geq \tilde{\rho}$, we necessarily have $|x_i - x_j| < \tilde{\rho}$. \square

H LocalRefineWS

We have already presented `LocalRefine` for which our performance analysis only deals with deterministic or independent input H_1 . `LocalRefineWS` can be seen as an extension of `LocalRefine` which allows to deal with non-independent random input H_1 . This extension will still use statistics like the l_{ij} and r_{ij} of `LocalRefine`, in order to determine comparisons $H_{ij}^{(\pi)}$. Before giving the algorithm, let us explain the issue when H_1 is non independent of A . `LocalRefine` uses statistics l_{ij}, r_{ij} that involve both the random data columns A_i, A_j , and the sets L_{ij}, R_{ij} computed from H_1 . So, when taking the input $H_1 = \widehat{H}_1$ in the meta algorithm `SALB`, we do not have independence between the sets L_{ij}, R_{ij} and the random columns

A_i, A_j . In order to avoid such statistical dependencies in the statistics l_{ij}, r_{ij} , we introduce `LocalRefineWS` which splits the data A to obtain the wanted independence. More precisely, it computes proxy sets pL_{ij}, pR_{ij} from the split data $A_{-(i,j)}$ where the i^{th} and j^{th} lines of A have been set to zero. Thus, pL_{ij}, pR_{ij} will be independent of the data columns A_i, A_j , and we will be able to analyze statistics like l_{ij}, r_{ij} on these sets pL_{ij}, pR_{ij} .

Unfortunately, there is a computational drawback for splitting data. The time complexity of SALB is $O(n^5)$ because of `LocalRefineWS`, whereas it would have been only $O(n^3)$ if `LocalRefine` were used (instead of `LocalRefineWS`). Indeed, `LocalRefineWS` repeats almost the entire procedure (i.e. steps like distance estimation – local bisection – test as in local refine) for each pair i, j left undetermined by `AgregLocalBisection`. The number of such pairs is at most $O(n^2)$, and the computational cost per iteration is $O(n^3)$. Each iteration involves a call to `DistanceEstimation` to obtain distance estimates $\widehat{D}_{-(i,j)}$ from $A_{-(i,j)}$, and then a call to `ProxyLocalBisection` to build proxy sets pL_{ij}, pR_{ij} from $\widehat{D}_{-(i,j)}$.

Algorithm 8 `LocalRefineWS`

Require: $(H_1, D, A, \sigma, \delta_1, \delta_2, \delta_3)$

Ensure: $H \in \{-1, 0, 1\}^{n \times n}$

```

1: Let  $S_{H_1} = \{(i, j) \in [n]^2 \text{ s.t. } i < j \text{ and } (H_1)_{ij} = 0\}$ 
2: for  $i \in [n]$  do
3:    $S_{H_1}^i = \{j \in [n] \text{ s.t. } (i, j) \in S_{H_1}\}$ 
4:   if  $S_{H_1}^i \neq \emptyset$  then
5:      $L_i = \{k : (H_1)_{ki} = -1\}$  and  $R_i = \{k : (H_1)_{ki} = 1\}$ 
6:     if  $L_i \neq \emptyset$  then
7:        $[pL_{ij}]_{j \in S_{H_1}^i} = \text{ProxySetDataSplit}(i, S_{H_1}^i, L_i, D, A, \delta_1, \delta_2, \delta_3)$ 
8:     else
9:        $pL_{ij} = \emptyset$  for all  $j \in S_{H_1}^i$ 
10:    end if
11:    if  $R_i \neq \emptyset$  then
12:       $[pR_{ij}]_{j \in S_{H_1}^i} = \text{ProxySetDataSplit}(i, S_{H_1}^i, R_i, D, A, \delta_1, \delta_2, \delta_3)$ 
13:    else
14:       $pR_{ij} = \emptyset$  for all  $j \in S_{H_1}^i$ 
15:    end if
16:  end if
17: end for
18: for  $(i, j) \in S_{H_1}$  do
19:    $H_{ij} = \text{Test}(i, j, pL_{ij}, pR_{ij}, A, \sigma)$  and  $H_{ji} = -H_{ij}$ .
20: end for

```

Given a comparison matrix $H_1 \in \{-1, 0, 1\}^{n \times n}$, let $S_{H_1} \subset [n]^2$ the set of pairs (i, j) left undetermined by H_1 ,

$$S_{H_1} = \{(i, j) \in [n]^2 : i < j \text{ and } (H_1)_{ij} = 0\} . \quad (81)$$

We will also use the notation

$$S_{H_1}^i = \{j \in [n] : (i, j) \in S_{H_1}\} \quad i \in [n] , \quad (82)$$

for denoting the indices j such that $(i, j) \in S_{H_1}$. The objective in `LocalRefineWS`(H_1, \dots) is to compute, for all $(i, j) \in S_{H_1}$, an estimate H_{ij} of $H_{ij}^{(\pi)}$. In lines 7 and 12, `LocalRefineWS` computes proxy sets pL_{ij}, pR_{ij} which will be used (instead of the old sets L_{ij}, R_{ij}) in the statistical test in line 19. This test is the same as the one used earlier in `LocalRefine`, and is encapsulated in Algorithm 7. The key point here, is that the input sets pL_{ij}, pR_{ij} are independent of the data columns A_i, A_j that will be tested, thus avoiding any complex statistical dependencies. The construction of proxies pL_{ij}, pR_{ij} is done by `ProxySetDataSplit`:

Algorithm 9 `ProxySetDataSplit`

Require: $(i, S, G, D, A, \delta_1, \delta_2, \delta_3)$

Ensure: $[G_{ij}]_{j \in S}$

- 1: Set $k_i \in \arg \min_{\{k \in G \text{ s.t. } D_{ik} \geq \delta_3\}} D_{ik}$.
 - 2: **for** $j \in S$ **do**
 - 3: $A_{-(i,j)}$ = matrix equal to A but with 0 on $i^{\text{th}}, j^{\text{th}}$ rows/columns
 - 4: $\widehat{D}_{-(i,j)}$ = `DistanceEstimation`($A_{-(i,j)}$)
 - 5: Let $(G_{k_i}^{(1)}, G_{k_i}^{(2)}) = \text{LocalBisection}(k_i, \widehat{D}_{-(i,j)}, \delta_1, \delta_2, \delta_3)$
 - 6: **if** $G_{k_i}^{(1)} \subset G$ **then**
 - 7: $G_{ij} = G_{k_i}^{(1)}$
 - 8: **else**
 - 9: $G_{ij} = G_{k_i}^{(2)}$.
 - 10: **end if**
 - 11: **end for**
-

Since i is missing in the split data $A_{-(i,j)}$, we need a proxy for i . The proxy k_i (defined in line 1) is chosen among all elements $k \in G$ such that it is at a small distance D_{ik} of i . But for technical reasons, this distance cannot be smaller than δ_3 . Then, `DistanceEstimation` computes (from $A_{-(i,j)}$) an estimate $\widehat{D}_{-(i,j)}$ of the distance matrix $D_{-(i,j)}^*$; here we used the notation $D_{-(i,j)}^*$ to denote the original distance matrix D^* with zeros on the i^{th} and j^{th} rows; accordingly, the estimator $\widehat{D}_{-(i,j)}$ is a symmetric matrix in $\mathbb{R}^{n \times n}$ with zeros on the i^{th} and j^{th} rows. Finally, `LocalBisection` computes (from $\widehat{D}_{-(i,j)}$) two sets $G_{k_i}^{(1)}, G_{k_i}^{(2)}$ that we expect to be on different sides of k_i (by Lemma D.2 and D.3). Among these two sets, we select (in lines 7 and 9) the one that is located in a set G of reference.

I Proof of Proposition C.8

`LocalRefineWS` (in line 19) and `LocalRefine` uses the same algorithm `Test`, so the proofs of Proposition C.8 and Proposition C.7 (in appendix D.4) are similar at a high level: It invokes Lemma D.6 to obtain guarantees for the output H_{ij} of `Test`, and the difficulty is to prove that the hypotheses of Lemma D.6 are fulfilled. More precisely, for an input $H_1 \in \{-1, 0, 1\}^{n \times n}$, and any pair (i, j) in the set S_{H_1} of (81), `LocalRefineWS` computes an estimate $H_{ij} = \text{Test}(i, j, pL_{ij}, pR_{ij}, A, \sigma)$. Using the convenient notation

$$B_{ij}^- = pL_{ij} \quad \text{and} \quad B_{ij}^+ = pR_{ij} \quad , \quad (83)$$

Lemma D.6 will give us the desired guarantees for H_{ij} , if we succeed in proving the three conditions below:

- (1) that the following distance bounds hold for a constant $\tilde{\rho} \in [0, r']$ and all $(i, j) \in S_{H_1}$,

$$|x_i - x_j| < \tilde{\rho} , \quad (84)$$

where $x_j = \pi_j/n$ for all $j \in [n]$.

- (2) that the following event holds with probability $1 - 8/n^2$,

$$\bigcap_{i,j \in S_{H_1}} \mathcal{E}_{B_{ij}^\pm} = \left\{ \max_{\substack{(i,j) \in S_{H_1} \\ \epsilon \in \{\pm\}}} \frac{1}{\sqrt{2\#B_{ij}^\epsilon}} \left| \sum_{\ell \in B_{ij}^\epsilon} (E_{i\ell} - E_{j\ell}) \right| < \sqrt{10 \log(n)} \right\} \quad (85)$$

- (3) that the following inclusions are satisfied for all $(i, j) \in S_{H_1}$, with respect to the same $\epsilon \in \{\pm\}$ as in the error bound (36) assumed on the input H_1 (of `LocalRefineWS`),

$$\begin{aligned} sL_{ij}^*(r'') &\subset B_{ij}^{-\epsilon} \subset L_{ij}^* && \text{when } x_i \wedge x_j \geq 1/2 - \tilde{\rho}/2 , \\ sR_{ij}^*(r'') &\subset B_{ij}^\epsilon \subset R_{ij}^* && \text{when } x_i \vee x_j \leq 1/2 + \tilde{\rho}/2 . \end{aligned} \quad (86)$$

By Lemma D.9, condition (84) is satisfied, with respect to the same $\tilde{\rho}$ as in the error bound (36) on H_1 . The high probability event (85) is proved in appendix I.1. The inclusions (86) are proved in in appendix I.2. Thus, all hypotheses of Lemma D.6 hold; this gives the wanted guarantees for the output H of `LocalRefineWS`. Proposition C.8 follows. \square

I.1 Proof for high probability event (85)

Complementing the notation $B_{ij}^- = pL_{ij}$ and $B_{ij}^+ = pR_{ij}$ introduced in (83), we will also write

$$B_i^- = L_i \quad \text{and} \quad B_i^+ = R_i . \quad (87)$$

With these notations and $\epsilon \in \{\pm\}$, we have that the sets B_{ij}^ϵ for $j \in S_{H_1}^i$ are defined in `LocalRefineWS` by

$$[B_{ij}^\epsilon]_{j \in S_{H_1}^i} = \text{ProxySetDataSplit}(i, S_{H_1}^i, B_i^\epsilon, D \dots) .$$

More precisely, following the lines of `ProxySetDataSplit` (Algorithm 9), this means that, for

$$k_i^\epsilon := k_i \in \underset{k \in B_i^\epsilon \text{ s.t. } D_{ik} \geq \delta_3}{\text{arg min}} D_{ik} , \quad (88)$$

and

$$(G_{k_i}^{(1)}, G_{k_i}^{(2)}) = \text{LocalBisection}(k_i, \widehat{D}_{-(i,j)} \dots) , \quad (89)$$

we have

$$B_{ij}^\epsilon = G_{k_i}^{(1)} \quad \text{or} \quad B_{ij}^\epsilon = G_{k_i}^{(2)} . \quad (90)$$

Thus, B_{ij}^ϵ depends on the possibly random quantities $k_i^\epsilon \in [n]$ and $\widehat{D}_{-(i,j)} \in \mathbb{R}^{n \times n}$, and some random variable $q_i \in [2]$ representing the two outcomes situation in (90) (that is: $q_i = 1$ when

$B_{ij}^\epsilon = G_{k_i}^{(1)}$, and $q_i = 2$ when $B_{ij}^\epsilon = G_{k_i}^{(2)}$). Therefore, there exists a deterministic function F (from the space product $[n] \times [2] \times \mathbb{R}^{n \times n}$ to the set of all subsets of $[n]$) such that

$$B_{ij}^\epsilon = F(k_i^\epsilon, q_i, \widehat{D}_{-(i,j)}) .$$

Consider deterministic integers $k \in [n]$ and $q \in [2]$. Since $\widehat{D}_{-(i,j)}$ is computed from $A_{-(i,j)}$ only, it is independent of A_i and A_j , and hence of E_i and E_j . Therefore, conditionally to $\widehat{D}_{-(i,j)}$, the set $F(k, q, \widehat{D}_{-(i,j)})$ for $k \in [n], q \in [2]$ is deterministic, and the $2(n-1)$ real numbers $\{E_{i\ell} : \ell \in [n], \ell \neq i\}$ and $\{E_{j\ell} : \ell \in [n], \ell \neq j\}$ are independent sub-Gaussian random variables, with zero means and variance proxies smaller than 1. Then, for fixed $k \in [n]$ and $q \in [2]$, we can apply a standard concentration inequality [Rigollet and Hütter, 2023, Corollary 1.7] and obtain

$$\mathbb{P} \left[\frac{1}{\sqrt{2 \#F(k, q, \widehat{D}_{-(i,j)})}} \left| \sum_{\ell \in F(k, q, \widehat{D}_{-(i,j)})} (E_{i\ell} - E_{j\ell}) \right| \geq t \mid \widehat{D}_{-(i,j)} \right] \leq 2e^{-t^2/2} ,$$

for all $t > 0$ (using the convention $0/0 = 0$ and $\sum_{k \in \emptyset} = 0$). Since this inequality holds for all $\widehat{D}_{-(i,j)}$ computed by `DistanceEstimation`, we have the same upper bound without conditioning (to $\widehat{D}_{-(i,j)}$). Taking $t = \sqrt{10 \log(n)}$, and a union bound over all $\epsilon \in \{\pm\}$, $k \in [n]$, $q \in [2]$ and $(i, j) \in S_{H_1}$, we obtain

$$\mathbb{P} \left[\max_{\substack{(i,j) \in S_{H_1} \\ \epsilon \in \{\pm\} \\ q \in [2], k \in [n]}} \frac{1}{\sqrt{2 \#F(k, q, \widehat{D}_{-(i,j)})}} \left| \sum_{\ell \in F(k, q, \widehat{D}_{-(i,j)})} (E_{i\ell} - E_{j\ell}) \right| \geq \sqrt{10 \log(n)} \right] \leq \frac{8}{n^2} .$$

This bound holds uniformly over all sets $F(k, q, \widehat{D}_{-(i,j)})$ for $k \in [n], q \in [2]$. In particular, it holds for the sets $B_{ij}^\epsilon = F(k_i^\epsilon, q_i, \widehat{D}_{-(i,j)})$. Thus, the event (85) occurs with probability $1 - 8/n^2$. \square

I.2 Proof for inclusions (86)

For convenience, we encapsulate in Lemma I.1 the result we want to prove. We recall the notation $x_j = \pi_j/n$ for all $j \in [n]$.

Lemma I.1. *For any $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r, \gamma, r'$ and $0 \leq \omega, r''$, the following holds for any $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$ and $D \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ (as defined in Assumption 4.1), and any $\delta_1, \delta_2, \delta_3, \tilde{\rho}$ fulfilling the constraints (115), in particular the following constraints: $\tilde{\rho} \in (0, 1)$ and $\tilde{\rho} \leq \rho$ for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$, and $\tilde{\rho}' + \rho \leq r''$ for*

$$\tilde{\rho}' = \frac{(\phi_1 \vee \phi_2) + 2\omega}{\tilde{\alpha}} \quad \text{where} \quad \phi_1 = \tilde{\beta}(n^{-1} + \tilde{\rho}) \quad \text{and} \quad \phi_2 = \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega)) .$$

If the input H_1 of `LocalRefineWS` has an accuracy, for an $\epsilon \in \{\pm\}$,

$$(H_1)_{k\ell} = \epsilon H_{k\ell}^{(\pi)} \quad \text{for all } k, \ell \in [n] \text{ where } (H_1)_{k\ell} \neq 0 \text{ or } |x_k - x_\ell| \geq \tilde{\rho} , \quad (91)$$

then, we have the inclusions (86) for all $(i, j) \in S_{H_1}$, with the same $\epsilon \in \{\pm\}$ as in (91).

Proof of Lemma I.1. Without the loss of generality, we focus on the case where $\epsilon = +$ in (91), so that we have

$$(H_1)_{k\ell} = H_{k\ell}^{(\pi)} , \quad \text{wherever } (H_1)_{k\ell} \neq 0 \text{ or } |x_k - x_\ell| \geq \tilde{\rho} . \quad (92)$$

For any $(i, j) \in S_{H_1}$, the definition of S_{H_1} in (81) yields $(H_1)_{ij} = 0$, and hence $|x_i - x_j| < \tilde{\rho}$. Thus, we have at least one of the two inequalities: $x_i \wedge x_j \geq 1/2 - \tilde{\rho}/2$ or $x_i \vee x_j \leq 1/2 + \tilde{\rho}/2$. By symmetry, we can focus on a single case, say $x_i \vee x_j \leq 1/2 + \tilde{\rho}/2$. Then, the inclusions (86) (to prove) become

$$sR_{ij}^*(r'') \subset pR_{ij} \subset R_{ij}^* , \quad (93)$$

where we used the notation $B_{ij}^+ = pR_{ij}$ introduced in (83). By definition, the sets pR_{ij} are defined by

$$[pR_{ij}]_{j \in S_{H_1}^i} = \text{ProxySetDataSplit}(i, S_{H_1}^i, R_i, D, A, \delta_1, \delta_2, \delta_3) . \quad (94)$$

Reading `ProxySetDataSplit` (Algorithm 9), this means that, for

$$k_i \in \underset{k \in R_i \text{ s.t. } D_{ik} \geq \delta_3}{\arg \min} D_{ik} , \quad (95)$$

and

$$(G_{k_i}^{(1)}, G_{k_i}^{(2)}) = \text{LocalBisection}(k_i, \widehat{D}_{-(i,j)} \dots) , \quad (96)$$

we have

$$\begin{aligned} pR_{ij} &= G_{k_i}^{(1)} && \text{if } G_{k_i}^{(1)} \subset R_i , \\ pR_{ij} &= G_{k_i}^{(2)} && \text{otherwise .} \end{aligned} \quad (97)$$

Lemma I.2 ensures the existence of k_i in (95), and also that the set R_i (where k_i lives) is correctly located. This lemma uses the error bound assumption (92) on H_1 , and the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity assumption on D . The proof is in appendix J.1.

Lemma I.2. *For any $n \geq 4$, $\tilde{\rho} \in (0, 1/4]$, $\delta_3 \leq r$ and $\delta_3 + \omega < \tilde{\alpha}/4$, the following holds for all $i \in [n]$. If $x_i \leq 1/2 + \tilde{\rho}/2$, and H_1 satisfies (92), and D is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$, then*

$$\emptyset \neq sR_i^*(\tilde{\rho}) \subset R_i \subset R_i^* ,$$

and $k_i \in R_i$ is well-defined.

Lemma I.3 gives a control on the distance $|x_{k_i} - x_i|$ between i and k_i . The proof is in appendix J.2.

Lemma I.3. *Define $\rho = (\delta_2 + \omega)/\tilde{\alpha}$ and $\phi_1 = \tilde{\beta}(n^{-1} + \tilde{\rho})$ and $\phi_2 = \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega))$ and $\tilde{\rho}' = \tilde{\alpha}^{-1}((\phi_1 \vee \phi_2) + 2\omega)$. Under the constraints $|x_i - x_j| < \tilde{\rho}$, and $\tilde{\rho} \leq \rho$, and*

$$\omega + \tilde{\beta}\rho < \delta_3 , \quad n^{-1} + \tilde{\rho} \leq r , \quad \phi_2 \leq \tilde{\beta}r , \quad (\phi_1 \vee \phi_2) + \omega \leq r ,$$

and under the hypotheses of Lemma I.2, we have the following

$$x_j \vee (x_i + \rho) < x_{k_i} < x_i + \tilde{\rho}' .$$

To have guarantees for the sets $(G_{k_i}^{(1)}, G_{k_i}^{(2)})$ in (96), we essentially need to check that $\widehat{D}_{-(i,j)}$ is a good enough estimator of $D_{-(i,j)}^*$. Since the i^{th} and j^{th} rows/columns of these matrices are null, we will describe the regularity of $\widehat{D}_{-(i,j)}$ via a direct extension of the \mathcal{LDE} assumption to matrices of smaller support $S \subset [n]$. Here, the support of $\widehat{D}_{-(i,j)}$ and $D_{-(i,j)}^*$ is denoted by $S_{-\{i,j\}} = \{\ell \in [n], \ell \neq i, j\}$. The adapted \mathcal{LDE} assumption to arbitrary support S is properly defined in Assumption J.1, and is denoted by $\mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, \omega, r, S)$. Lemma I.4 ensures that, with high probability, the $\widehat{D}_{-(i,j)}$ are simultaneously in $\mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, 2\omega_n, r, S_{-\{i,j\}})$ for $(i, j) \in S_{H_1}$. We introduce more formally this event: Given any $\tilde{\alpha}, \tilde{\beta}, \omega, r > 0$, let

$$\bigcap_{(i,j) \in S_{H_1}} \mathcal{F}_{ij}(\tilde{\alpha}, \tilde{\beta}, \omega) = \left\{ \text{for all } (i, j) \in S_{H_1} : \widehat{D}_{-(i,j)} \text{ is } \mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, \omega, r, S_{-\{i,j\}}) \right\}. \quad (98)$$

Lemma I.4. *For any $\tilde{\alpha}, \tilde{\beta}, r > 0$, the following holds when the distance matrix D^* is $\mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, 0, r)$. Then event $\bigcap_{(i,j) \in S_{H_1}} \mathcal{F}_{ij}(\tilde{\alpha}, \tilde{\beta}, 2\omega_n)$ in (98) occurs with probability $1 - 1/n^2$, where ω_n is the distance estimation error defined in (32).*

The proof of Lemma I.4 is in appendix J.3. We can now apply Lemma I.5; it will give the desired inclusions (93) for the output pR_{ij} in (97). The proof of Lemma I.5 is in appendix J.4.

Lemma I.5. *For any $i, j \in [n]$, $i \neq j$, any $\tilde{\rho} \in (0, 1)$, any $\delta_1, \delta_2, \delta_3, n$ complying with the constraints (44) and also the constraints*

$$\delta_3 \leq r, \quad \delta_3 + \omega < \tilde{\alpha}/4, \quad \tilde{\rho} + 2t' \leq 1/2,$$

the following holds. If $x_i \leq 1/2 + \tilde{\rho}/2$, and k_i and R_i satisfy, for some $t' > t > 0$,

$$x_j \vee (x_i + t) < x_{k_i} < x_i + t' \quad \text{and} \quad sR_i^*(t) \subset R_i \subset R_i^*, \quad (99)$$

and $\widehat{D}_{-(i,j)}$ is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r, S_{-\{i,j\}}]$ (as in Assumption J.1), then, the output pR_{ij} satisfies the following inclusions, for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$,

$$sR_{ij}^*(t' + \rho) \subset pR_{ij} \subset R_{ij}^*. \quad (100)$$

By Lemma I.2, I.3 and I.4, we can take $t = \rho$ (with $\rho \geq \tilde{\rho}$), and $t' = \tilde{\rho}'$ (with $\tilde{\rho}'$ defined in Lemma I.3) and $\omega = 2\omega_n$, so that, all conditions of Lemma I.5 are satisfied. Then, for $\tilde{\rho}' + \rho \leq r''$, the conclusion of Lemma I.5 gives the inclusions (93). The proof of Lemma I.1 is complete.

J Proofs of Lemma I.2 to I.5

J.1 Proof of Lemma I.2

We recall that R_i is defined by LocalRefineWS by $R_i = \{k : (H_1)_{ik} = -1\}$. The accuracy (92) of H_1 ensures that R_i is on the right side of i , and it includes all objects at distance (at least) $\tilde{\rho}$ on that side. This gives

$$sR_i^*(\tilde{\rho}) \subset R_i \subset R_i^*,$$

where we recall the notations $sR_i^*(\tilde{\rho}) = \{\ell : x_\ell > x_i + \tilde{\rho}\}$ and $R_i^* = \{\ell : x_\ell > x_i\}$.

To check that $sR_i^*(\tilde{\rho}) \neq \emptyset$ for $\tilde{\rho} \leq 1/4$, one observes that $x_i \leq 5/8$ (since $x_i \leq 1/2 + \tilde{\rho}/2$ by assumption), and the $x_j = \pi_j/n$ for $j \in [n]$ are evenly spread in $[0, 1]$, with a spacing $1/n \leq 1/4$. Thus, we have shown that

$$\emptyset \neq sR_i^*(\tilde{\rho}) \subset R_i \subset R_i^* . \quad (101)$$

To complement the proof of Lemma I.2, it remains to check that k_i is well-defined. By definition (95), we have $k_i \in \arg \min_{\mathcal{K}_i} D_{ik}$ where

$$\mathcal{K}_i = \{k \in R_i : D_{ik} \geq \delta_3\} . \quad (102)$$

Thus, it suffices to show that $\mathcal{K}_i \neq \emptyset$. Assuming (w.l.o.g.) $\pi = id$ (i.e. $x_k = k/n$ for all $k \in [n]$), we will show that $n \in \mathcal{K}_i$. This is equivalent to show that $n \in R_i$ and $D_{in} \geq \delta_3$. The property (101) yields $n \in R_i$. To prove $D_{in} \geq \delta_3$, we proceed by contradiction: if $D_{in} < \delta_3$, the \mathcal{LDE} regularity (with $\delta_3 \leq r$) would give

$$x_n - x_i < (\delta_3 + \omega)/\tilde{\alpha} < 1/4 ,$$

since $\delta_3 + \omega < \tilde{\alpha}/4$. But, for $\tilde{\rho} \leq 1/2$, the assumption $x_i \leq 1/2 + \tilde{\rho}/2$ yields

$$x_n - x_i \geq 1 - (1/2 + \tilde{\rho}/2) \geq 1/4 ,$$

hence a contradiction between the last two displays. Therefore, we have $D_{in} \geq \delta_3$. Combining with $n \in R_i$, we obtain $n \in \mathcal{K}_i$. Thus $\mathcal{K}_i \neq \emptyset$ and the index k_i is well defined. The proof of Lemma I.2 is complete.

J.2 Proof of Lemma I.3

Without the loss of generality, we assume that $\pi = id$, so that $x_\ell = \ell/n$ for all $\ell \in [n]$.

Lower bound: We recall that $k_i \in \arg \min_{\mathcal{K}_i} D_{ik}$ where the set \mathcal{K}_i is defined in (102). This means that $k_i \in \mathcal{K}_i \subset R_i$. Combining with the the inclusion $R_i \subset R_i^*$ of Lemma I.2, we obtain $k_i \in R_i^*$, which implies $x_i < x_{k_i}$. Then, the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity and the inequality $D_{ik_i} \geq \delta_3$ give

$$x_{k_i} - x_i \geq \frac{\delta_3 - \omega}{\tilde{\beta}} > \rho ,$$

since $\delta_3 > \omega + \tilde{\beta}\rho$. Thus $x_i + \rho < x_{k_i}$ which is half of the lower bound of Lemma I.3. Using the assumption $|x_i - x_j| < \tilde{\rho}$, we have

$$x_j = (x_j - x_i) + (x_i - x_{k_i}) + x_{k_i} < \tilde{\rho} - \rho + x_{k_i} \leq x_{k_i} ,$$

since $\tilde{\rho} \leq \rho$. This yields $x_j < x_{k_i}$, the other half of the lower bound of Lemma I.3. \square

Upper bound: Given the set $\mathcal{K}_i = \{k \in R_i : D_{ik} \geq \delta_3\}$ from (102), we define b_i^+ as an element of \mathcal{K}_i which has the lowest position in the ordering π , that is:

$$b_i^+ = \min \mathcal{K}_i . \quad (103)$$

Note that b_i^+ is well defined since $\emptyset \neq \mathcal{K}_i \subset [n]$ (where the non-emptiness was checked in the proof of Lemma I.2). We also have $b_i^+ - 1 \in [n] \setminus \mathcal{K}_i$. (Indeed, by definition of b_i , we have $b_i^+ - 1 \notin \mathcal{K}_i$, and, to see that $b_i^+ - 1 \in [n]$, one can observe that $i + 1 \leq b_i^+$ (since $b_i^+ \in \mathcal{K}_i \subset R_i \subset R_i^*$, where the second inclusion is guaranteed by Lemma I.2).) By definition of \mathcal{K}_i , the fact $b_i^+ - 1 \in [n] \setminus \mathcal{K}_i$ means that $b_i^+ - 1 \notin R_i$ or $D_{i, b_i^+ - 1} < \delta_3$. Both cases are analyzed separately.

◦ *Case $b_i^+ - 1 \notin R_i$* : Since $sR_i^*(\tilde{\rho}) \subset R_i$ by Lemma I.2, we have $x_{b_i^+ - 1} \notin sR_i^*(\tilde{\rho})$, which means $x_{b_i^+ - 1} \leq x_i + \tilde{\rho}$. We also have $x_i < x_{b_i^+}$, since $b_i^+ \in \mathcal{K}_i \subset R_i \subset R_i^*$ (by Lemma I.2). Combining these inequalities, we obtain

$$0 < x_{b_i^+} - x_i = (x_{b_i^+} - x_{b_i^+ - 1}) + (x_{b_i^+ - 1} - x_i) \leq n^{-1} + \tilde{\rho} . \quad (104)$$

By definition of k_i , we have $D_{k_i i} \leq D_{b_i^+ i}$. Then, the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity of D (for $n^{-1} + \tilde{\rho} \leq r$) yields

$$D_{k_i i} \leq D_{b_i^+ i} \leq \tilde{\beta}(n^{-1} + \tilde{\rho}) + \omega = \phi_1 + \omega , \quad (105)$$

where we wrote $\phi_1 = \tilde{\beta}(n^{-1} + \tilde{\rho})$.

◦ *Case $D_{ib_i^+ - 1} < \delta_3$* : The distance matrix D^* satisfies the triangular inequality, so we could have assumed that the input D satisfies the triangular inequality too. Then, we would directly obtain $D_{k_i i} \leq D_{ib_i^+} \leq D_{i, b_i^+ - 1} + D_{b_i^+ - 1, b_i^+} < \delta_3 + \beta n^{-1} + \omega$, using the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity of D (for $n^{-1} \leq r$). Nevertheless, we do not need the triangular inequality for our analysis; below, we give a proof without using it.

The $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity (for $\delta_3 \leq r$) yields $\tilde{\alpha}(x_{b_i^+ - 1} - x_i) < \delta_3 + \omega$. Then, with the same decomposition as in (104), we obtain

$$|x_{b_i^+} - x_i| < n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega) .$$

We apply the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity again (for $n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega) \leq r$), and derive

$$D_{ib_i^+} < \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega)) + \omega .$$

Using the definition of k_i , we get

$$D_{k_i i} \leq D_{b_i^+ i} < \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega)) + \omega := \phi_2 + \omega , \quad (106)$$

where we wrote $\phi_2 = \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \omega))$.

◦ *Conclusion* : Combining the bounds (105-106) and the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity (for $(\phi_1 \vee \phi_2) + \omega \leq r$) we obtain

$$\tilde{\alpha}|x_{k_i} - x_i| \leq D_{k_i i} + \omega < (\phi_1 \vee \phi_2) + 2\omega .$$

This gives

$$x_{k_i} < x_i + \tilde{\rho}' , \quad \text{where } \tilde{\rho}' = \tilde{\alpha}^{-1}((\phi_1 \vee \phi_2) + 2\omega) ,$$

which is the upper bound of Lemma I.3. The proof of the lemma is complete.

J.3 Proof of Lemma I.4

The next assumption is a direct extension of the $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ regularity to matrices of arbitrary support $S \subset [n]$. Here, we use the notation $x_i = \pi_i/n$ for all i .

Assumption J.1 (Local Distance Equivalence with support S). *For any constants $0 < \tilde{\alpha} \leq \tilde{\beta}$, $0 < r$, $0 \leq \omega$ and a set $S \subset [n]$, let $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r, S]$ be the collection of symmetric matrices D in $\mathbb{R}^{n \times n}$ that satisfy*

$$\tilde{\alpha}|x_i - x_j| - \tilde{\omega} \leq D_{ij} \leq \tilde{\beta}|x_i - x_j| + \tilde{\omega} \quad (107)$$

for all $i, j \in S$ such that $|x_i - x_j| \wedge D_{ij} \leq r$.

We have already seen that, when D^* is $\mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, 0, r)$, the estimator \hat{D} is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega_n, r]$ with probability $1 - 1/n^4$ (by Lemma C.5). To extend these guarantees to $D_{-(i,j)}^*$ and $\hat{D}_{-(i,j)}$, we will apply Lemma L.5, which is a generalization of Proposition C.5 for general positions $x_i \in [0, 1]$ (instead of the regular grid of $[0, 1]$).

Fix $(i, j) \in [n]^2, i \neq j$. Since D^* is $\mathcal{LDE}(\tilde{\alpha}, \tilde{\beta}, 0, r)$, we have that $D_{-(i,j)}^*$ is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r, S_{-\{i,j\}}]$ as in Assumption J.1. The $(n - 2)$ latent positions x_k , for $k \neq i, j$, satisfy the spreading condition (18) for $\eta = 3/n$. (In fact, this value $\eta = 3/n$ is attained by the configuration of x_k 's where the removed points x_i, x_j are consecutive points in the ordering π). Then, Lemma L.5 (with a sample size $n - 2$) ensures that, with probability $1 - 1/n^4$, the matrix $\hat{D}_{-(i,j)}$ is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega_{n, \frac{3}{n}}, r, S_{-\{i,j\}}]$, where the distance estimator error $\omega_{n, \frac{3}{n}}$ is given by

$$\omega_{n, \frac{3}{n}} = C \left(\sqrt{\frac{3\tilde{\beta}}{n}} + \sqrt{(\sigma + 1)\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \right).$$

Since $\sqrt{3} \leq 2$, we have $\omega_{n, \frac{3}{n}} \leq 2\omega_n$. Thus $\hat{D}_{-(i,j)}$ is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 2\omega_n, r, S_{-\{i,j\}}]$.

Taking a union bound over all pairs $i, j \in [n]$, we obtain that, with probability $1 - 1/n^2$, all $\hat{D}_{-(i,j)}$ are simultaneously in $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 2\omega_n, r, S_{-\{i,j\}}]$ for all i, j . Lemma I.4 follows.

J.4 Proof of Lemma I.5

Assume (w.l.o.g.) $\pi = id$, so that $x_\ell = \ell/n$ for all $\ell \in [n]$. Recall the definition (97) of the output pR_{ij} : we first compute two sets $(G_{k_i}^{(1)}, G_{k_i}^{(2)}) = \text{LocalBisection}(k_i, \hat{D}_{-(i,j)} \dots)$, and then the output is equal to $pR_{ij} = G_{k_i}^{(1)}$ if $G_{k_i}^{(1)} \subset R_i$, and to $pR_{ij} = G_{k_i}^{(2)}$ otherwise. Using the convenient notations $k = k_i$ and $\tilde{D} = \hat{D}_{-(i,j)}$ for the rest of the proof, we momentarily assume that $\tilde{D}_{kn} \geq \delta_3$ (we will prove it at the end of the section). By construction of LocalBisection (and Lemma F.6), the inequality $\tilde{D}_{kn} \geq \delta_3$ yields $n \in G_k^{(1)}$ or $n \in G_k^{(2)}$. We study both cases separately.

◦ **Case $n \in G_k^{(1)}$.** We want to apply Lemma D.3 to have guarantees for the output $G_k^{(1)}$ of LocalBisection . The conditions of Lemma D.3 are indeed fulfilled, since the $\delta_1, \delta_2, \delta_3$ comply with the constraints (44) and \tilde{D} is $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r, S_{-(i,j)}]$. The fact that \tilde{D} is only \mathcal{LDE} w.r.t. $S_{-(i,j)}$ (instead of the whole set $[n]$) is not an issue here, because we only focus

on the output $G_k^{(1)}$ for $k \neq i, j$. Then, Lemma D.3 and $n \in G_k^{(1)}$ together yield

$$sR_k^*(\rho) \subset G_k^{(1)} \subset R_k^* , \quad (108)$$

for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$. On the one hand, the assumption (99) gives $x_j \vee x_i < x_k$, that is, $k \in R_{ij}^*$, and hence $R_k^* \subset R_{ij}^*$. Combining with (108), we obtain

$$G_k^{(1)} \subset R_{ij}^* . \quad (109)$$

On the other hand, the assumption (99) also gives $x_i < x_k < x_i + t'$, which yields $sR_i^*(t' + \rho) \subset sR_k^*(\rho)$. Then, using the trivial inclusion $sR_{ij}^*(t' + \rho) \subset sR_i^*(t' + \rho)$ and (108), we obtain

$$sR_{ij}^*(t' + \rho) \subset G_k^{(1)} . \quad (110)$$

Putting everything together, we get

$$sR_{ij}^*(t' + \rho) \subset G_k^{(1)} \subset R_{ij}^* . \quad (111)$$

Therefore, it remains to prove that $pR_{ij} = G_k^{(1)}$. By construction of `ProxySetDataSplit`, this equality will occur iff $G_k^{(1)} \subset R_i$. The assumption (99) gives $x_i + t < x_k$ and $sR_i^*(t) \subset R_i$, therefore we have $k \in sR_i^*(t) \subset R_i$. In particular, this implies that any set on the right side of k is in R_i . Therefore, the inclusion $G_k^{(1)} \subset R_k^*$ of (108) yields $G_k^{(1)} \subset R_i$. This means that $pR_{ij} = G_k^{(1)}$ (by construction of `ProxySetDataSplit`). Substituting this into (111), we obtain the conclusion of Lemma I.5. \square

◦ **Case $n \in G_k^{(2)}$.** We are back to the situation above, and we can proceed in a similar fashion to obtain the inclusions (111) with $G_k^{(2)}$ instead of $G_k^{(1)}$. Then, we have to check the equality $G_k^{(2)} = pR_{ij}$. By definition of `ProxySetDataSplit`, this equality holds iff $G_k^{(1)} \not\subset R_i$. On the one hand, the assumption (99) gives $R_i \subset R_i^* \subset R_{ij}^*$, where the last inclusion is trivial. Therefore $1 \notin R_i$.

On the other hand, Lemma D.3 ensures that $G_k^{(1)}, G_k^{(2)}$ are on different sides of k . Then, $G_k^{(2)}$ is necessarily on the right side of k (since $n \in G_k^{(2)}$) and hence $G_k^{(1)}$ is on the left side. Using Lemma D.3, we obtain that $G_k^{(1)}$ is non-empty and contains all objects at distance ρ on the left side of k . This yields $1 \in G_k^{(1)}$.

Combining the conclusions of the last two paragraphs, we obtain $G_k^{(1)} \not\subset R_i$. Thus $G_k^{(2)} = pR_{ij}$. The proof for the case $n \in G_k^{(2)}$ is complete. \square

◦ **Proof of the initial assumption $\tilde{D}_{kn} \geq \delta_3$.** By contradiction, if $\tilde{D}_{kn} < \delta_3$, the \mathcal{LDE} regularity (for $\delta_3 \leq r$) would give $x_n - x_k < (\delta_3 + \omega)/\tilde{\alpha} < 1/4$.

On the other hand, we have $0 < x_k - x_i < t'$ by assumption (99). Combining with the assumption $x_i \leq 1/2 + \tilde{\rho}/2$, we obtain the following contradiction

$$x_n - x_k = x_n - x_i + (x_i - x_k) > 1 - (1/2 + \tilde{\rho}/2) - t' = 1/2 - \tilde{\rho}/2 - t' \geq 1/4 ,$$

using $x_n = 1$ in the first inequality, and $\tilde{\rho} + 2t' \leq 1/2$ in the last inequality.

Thus, we have $\tilde{D}_{kn} \geq \delta_3$. \square

Lemma I.5 is proved.

K Conditions on the tuning parameters

The rate in Theorem 2.1 still holds when the constants α, β, σ depend on n , but the conditions on the tuning parameters $\delta_1, \delta_2, \delta_3$ are more intricate. To state these general conditions on $\delta_1, \delta_2, \delta_3$, we first need to recall some key quantities. For $F \in \mathcal{BL}[\alpha, \beta]$, recall that Proposition C.5 gives the following bound on the estimation error of \widehat{D} :

$$\omega_n = C \left(\sqrt{\frac{\beta}{n}} + \sqrt{(\sigma + 1)\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \right)$$

where C is a numerical constant. Then, taking the input $D = \widehat{D}$ in `AgregLocalBisection`, the output \widehat{H}_1 of `AgregLocalBisection` has an estimation error bounded by $\rho = (\delta_2 + \omega_n)/\alpha$ (Proposition C.6). With the definition of these quantities ω_n and ρ , we are now ready to give the conditions on $\delta_1, \delta_2, \delta_3$ for which the rate in Theorem 2.1 is achieved:

$$\delta_1 \geq 2\omega_n + \frac{\beta}{n}, \quad \delta_2 > 2\omega_n + \frac{2\beta}{\alpha}(\delta_1 + 2\omega_n), \quad \frac{1}{4} > \frac{n^{-1} \vee (\delta_1 + 2\omega_n) \vee (\delta_2 + 2\omega_n)}{1 \wedge (\alpha/2)}, \quad (112)$$

$$n^{-1} \leq \rho \leq 1/8, \quad \left(\frac{1}{4} \wedge (\alpha/16) \right) - 2\omega_n > \delta_3 > 2\omega_n + \beta\rho,$$

and also:

$$n^{-1} + \rho < \frac{1}{8}, \quad 4 \frac{(\phi_1 \vee \phi_2) + 2\omega_n}{\alpha} < \frac{1}{10} - \rho, \quad (113)$$

$$\phi_2 < \frac{\beta}{4}, \quad (\phi_1 \vee \phi_2) + 4\omega_n < \frac{1}{4},$$

where we wrote $\phi_1 = \beta(n^{-1} + \rho)$, and $\phi_2 = \beta(n^{-1} + 2\alpha^{-1}(\delta_3 + 2\omega_n))$. To sum up, if $\delta_1, \delta_2, \delta_3$ satisfy (112-113), the conclusion of Theorem 2.1 is still valid, even if α, β, σ depend on n . For convenience, we have summarized these conditions in a single line in (28). In appendix K.1, we briefly give the origins of the conditions (112-113) in our analysis.

K.1 Origins of the conditions

AgregLocalBisection. The conditions in (112) are used for the performance analysis of `AgregLocalBisection`. More precisely, assuming that the input D of `AgregLocalBisection` is in $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ for positive constants $\tilde{\alpha}, \tilde{\beta}, \omega, r$, we get guarantees for the output of `AgregLocalBisection` by using Lemma D.3 D.4, D.5 which require the following conditions:

$$\delta_1 \geq \omega + \frac{\tilde{\beta}}{n}, \quad \delta_2 > \omega + \frac{\tilde{\beta}}{\tilde{\alpha}}(\delta_1 + \omega), \quad r \geq \frac{n^{-1} \vee (\delta_1 + \omega) \vee (\delta_2 + \omega)}{1 \wedge \tilde{\alpha}}, \quad (114)$$

$$n^{-1} \leq \rho \leq 1/8, \quad (r \wedge (\tilde{\alpha}/8)) - \omega \geq \delta_3 > \omega + \tilde{\beta}\rho.$$

To obtain the conditions (112), we first check that the input $D = \widehat{D}$ is in $\mathcal{LDE}[\alpha/2, \beta, \omega_n, r]$ (see next paragraph), and then we simply substitute $\tilde{\alpha} = \alpha/2$, $\tilde{\beta} = \beta$ and $\omega = \omega_n$ in the

conditions (114). In fact, there is an extra factor 2 in front of the ω_n in the conditions (112) which comes from the conditions below associated to `LocalRefineWS` performance.

Let us check that \widehat{D} is in $\mathcal{LDE}[\alpha/2, \beta, \omega_n, r]$ when $F \in \mathcal{BL}[\alpha, \beta]$. Proposition C.5 ensures that, when $D^* \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, 0, r]$, the estimator \widehat{D} is in $\mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega_n, r]$. Besides, for $F \in \mathcal{BL}[\alpha, \beta]$, Lemma C.1 ensures that D^* is in $\mathcal{LDE}[\alpha/2, \beta, 0, r]$. Thus we have $\widehat{D} \in \mathcal{LDE}[\alpha/2, \tilde{\beta}, \omega_n, r]$.

LocalRefineWS. SALB also calls to `LocalRefineWS` whose main input is a comparison matrix H_1 . Proposition C.8 requires that the error of H_1 is less than a constant $\tilde{\rho}$. This is guaranteed for $H_1 = \widehat{H}_1$ by Proposition C.6 which gives the error bound $\tilde{\rho} = \rho = (\delta_2 + \omega)/\tilde{\alpha} = 2(\delta_2 + \omega_n)/\alpha$ when \widehat{D} is in $\mathcal{LDE}[\alpha/2, \beta, \omega_n, r]$ (or $F \in \mathcal{BL}[\alpha, \beta]$). Besides, the performance analysis of `LocalRefineWS` also requires the conditions (114) because `LocalRefineWS` uses parts of `AggregLocalBisection`. But, `LocalRefineWS` uses both $D \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$ and estimates $\widehat{D}_{-(i,j)}$ (from the split data $A_{-(i,j)}$) in $\mathcal{LDE}[\alpha/2, \beta, 2\omega_n, r]$ (Lemma I.4), so the conditions (114) must be satisfied by $\omega \vee 2\omega_n$ (instead of just ω). In addition, the analysis of `LocalRefineWS` requires other conditions written in (115). For F as in Assumption 4.2 with constants $\gamma, r', r'' > 0$, and for an input $D \in \mathcal{LDE}[\tilde{\alpha}, \tilde{\beta}, \omega, r]$, we use the following conditions in the proof of Proposition C.8:

Conditions (114) with the replacement $\omega \leftarrow (\omega \vee 2\omega_n)$;

$$\tilde{\rho} \in [0, r'], \quad \tilde{\rho} \leq \rho, \quad \tilde{\rho}' + \rho \leq r'', \quad \tilde{\rho} + 2\tilde{\rho}' \leq 1/2 ; \quad (115)$$

$$\tilde{\omega} + \tilde{\beta}\tilde{\rho} < \delta_3 , \quad n^{-1} + \tilde{\rho} \leq r , \quad \phi_2 \leq \tilde{\beta}r , \quad (\phi_1 \vee \phi_2) + \tilde{\omega} \leq r ;$$

for the notation $\tilde{\omega} = \omega \vee 2\omega_n$ and $\rho = (\delta_2 + \tilde{\omega})/\alpha$, and

$$\tilde{\rho}' = \frac{(\phi_1 \vee \phi_2) + 2\tilde{\omega}}{\tilde{\alpha}} \quad \text{with} \quad \phi_1 = \tilde{\beta}(n^{-1} + \tilde{\rho}) \quad \text{and} \quad \phi_2 = \tilde{\beta}(n^{-1} + \tilde{\alpha}^{-1}(\delta_3 + \tilde{\omega})) .$$

The conditions (115) yield conditions that are satisfied under the constraints (113) and (112). Indeed, we have $\tilde{\rho} = \rho$ by Proposition C.6. Also, for $F \in \mathcal{BL}[\alpha, \beta]$, Lemma C.2 ensures that F fulfills Assumption 4.2 for $\gamma = \alpha/4$ and any $r' \in (0, 1/5)$ and $r'' \in [0, 1/10)$. Besides, for $F \in \mathcal{BL}[\alpha, \beta]$, Lemma C.1 guarantees that D^* that is $\mathcal{LDE}[\alpha/2, \beta, 0, r]$ for any $r \in (0, 1/4)$. Then, Proposition C.5 gives $\widehat{D} \in \mathcal{LDE}[\alpha/2, \beta, \omega_n, r]$. Thus, for $D = \widehat{D}$ we have $\omega = \omega_n$, and hence $\tilde{\omega} = \omega \vee 2\omega_n = 2\omega_n$. Substituting in (115) the constants $\tilde{\alpha} = \alpha/2$, $\tilde{\beta} = \beta$, $\tilde{\omega} = 2\omega_n$, and taking the largest possible values for $r \in (0, 1/4)$, $r' \in (0, 1/5)$ and $r'' \in [0, 1/10)$, we obtain conditions that are satisfied under the constraints (113) and (112).

L Proof of Theorem 5.1

L.1 General statement of the theorem

The next result extends Theorem 5.1 to the general situation where the α, β, σ may depend on n , and the maximum spacing η can be any constant that satisfies

$$\eta \leq c_{\alpha\beta} , \quad (116)$$

where $c_{\alpha\beta}$ is a positive constant only depending on α, β . In this general situation, the choices of tuning parameters $\delta_1, \delta_2, \delta_3$ may be summarized by the conditions

$$\beta\eta + C\sqrt{\beta\eta} + C_{\beta\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \leq \delta_1, \quad C_{\alpha\beta} \delta_1 \leq \delta_2 \leq C_\alpha, \quad C'_{\alpha\beta} \delta_2 \leq \delta_3 \leq C''_{\alpha\beta}, \quad (117)$$

where $C > 0$ is a numerical constant, and $C_{\beta\sigma}, C_{\alpha\beta}, C_\alpha, C'_{\alpha\beta}, C''_{\alpha\beta}$ are positive constants only depending on α, β, σ . The exact conditions (with explicit constants) for the choices of $\delta_1, \delta_2, \delta_3$ can be found in (131) (taking $\omega = \omega_n$ and $\tilde{\rho} = \rho$ in (131)).

Theorem L.1. *For $n \geq 8$, and $\alpha, \beta, \sigma > 0$, there exist positive constants $c_{\alpha\beta}, C_{\beta\sigma}, C_{\alpha\beta}, C_\alpha, C'_{\alpha\beta}, C''_{\alpha\beta}$ such that the following holds for any η as in (116), any $\delta_1, \delta_2, \delta_3$ fulfilling (117), for any $\tilde{\eta} > 0$ and $y_1, \dots, y_n \in [0, 1]$ complying with (18-19), and for any $F \in \mathcal{LBC}[(y_i), \alpha, \beta]$. With probability higher than $1 - 9/n^2$, the permutation $\hat{\pi}_o$ computed by SALB satisfies*

$$L_{\text{comp}}(\hat{\pi}_o, \pi) \leq C' \frac{\sigma}{\tilde{\eta}\alpha} \sqrt{\frac{\log(n)}{n}}, \quad (118)$$

where $C' > 0$ is a numerical constant.

L.2 Proof of Theorem 5.1 and L.1

The proof follows the same lines as for Theorem 2.1 and 4.1. Using the notation $x_i = y_{\pi_i}$ for all i , we first introduce the latent space extensions of Assumption 4.1 and 4.2.

Assumption L.1 (Latent Local Distance Equivalence). *For any constants $0 < \tilde{\alpha} \leq \tilde{\beta}$, $0 < r$, $0 \leq \omega$ and latent points $y_1 < \dots < y_n$ in $[0, 1]$, let $\mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$ be the collection of symmetric matrices $D \in \mathbb{R}^{n \times n}$ that satisfy*

$$\tilde{\alpha}|y_{\pi_i} - y_{\pi_j}| - \omega \leq D_{ij} \leq \tilde{\beta}|y_{\pi_i} - y_{\pi_j}| + \omega \quad (119)$$

for all $i, j \in [n]$ such that $|y_{\pi_i} - y_{\pi_j}| \wedge D_{ij} \leq r$.

Assumption L.2 (Latent Separated Cumulative similarities). *For any constants $0 < \gamma, r', r''$ and latent points $y_1 < \dots < y_n$ in $[0, 1]$, let $\mathcal{SCA}[(y_i), \gamma, r', r'']$ be the collection of matrices that satisfy, for all $i, j \in [n]$, $y_i < y_j$ and $|y_i - y_j| \leq r'$,*

$$\begin{aligned} \sum_{\ell: y_\ell < y_i - r''} F_{i\ell} - F_{j\ell} &\geq \gamma n |y_i - y_j| && \text{if } y_i \geq \frac{1 - r'}{2} \\ \sum_{\ell: y_\ell > y_j + r''} F_{j\ell} - F_{i\ell} &\geq \gamma n |y_i - y_j| && \text{if } y_j \leq \frac{1 + r'}{2}. \end{aligned} \quad (120)$$

Assumption L.1 and L.2 simply extend Assumption 4.1 and 4.2 in that they replace the regular grid $y_i = i/n$ with general points $y_i \in [0, 1]$.

The next result ensures that $\hat{\pi}_o$ achieves the rate $\sqrt{\log(n)/n}$ under Assumption L.1 and L.2, and for y_1, \dots, y_n fulfilling the conditions (18-19). The choice of tuning parameters $\delta_1, \delta_2, \delta_3$ is characterized by the following conditions:

$$\eta \leq c_{\tilde{\alpha}\tilde{\beta}rr'r''} \quad (121)$$

$$\tilde{\beta}\eta + C\sqrt{\tilde{\beta}\eta} + C_{\tilde{\beta}\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \leq \delta_1, \quad C_{\tilde{\alpha}\tilde{\beta}} \delta_1 \leq \delta_2 \leq C_{\tilde{\alpha}rr'r''}, \quad C'_{\tilde{\alpha}\tilde{\beta}} \delta_2 \leq \delta_3 \leq C_{\tilde{\alpha}\tilde{\beta}rr'r''},$$

where $C > 0$ is a numerical constant, and $c_{\tilde{\alpha}\tilde{\beta}rr'r''}, C_{\tilde{\beta}\sigma}, C_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}rr'r''}, C'_{\tilde{\alpha}\tilde{\beta}}, C_{\tilde{\alpha}\tilde{\beta}rr'r''}$ are positive constants only depending on $\tilde{\alpha}, \tilde{\beta}, \sigma, r, r', r''$.

Theorem L.2. *For $n \geq 8$ and $\tilde{\alpha}, \tilde{\beta}, \sigma, r, r', r'' > 0$, the following holds for any $\eta, \delta_1, \delta_2, \delta_3$ fulfilling (121), and $y_1, \dots, y_n \in [0, 1]$ complying with (18), and for any $F \in \mathcal{SCA}[(y_i), \gamma, r', r'']$ such that $D^* \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$. With probability higher than $1 - 9/n^2$, the permutation $\hat{\pi}_o$ computed by SALB satisfies*

$$L_{\text{comp}}(\hat{\pi}_o, \pi) \leq C' \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}, \quad (122)$$

where $C' > 0$ is a numerical constant.

We are now ready to prove Theorem 5.1 and its general version Theorem L.1.

Proof. (Theorem 5.1 and L.1) The next two lemmas show that the latent bi-Lipschitz assumption $F \in \mathcal{LBC}[(y_i), \alpha, \beta]$ implies Assumption L.1 and L.2.

Lemma L.3. *If F belongs to the class $\mathcal{LBC}[(y_i), \alpha, \beta]$ of latent bi-lipschitz matrices (as in Assumption 5.1) for $n \geq 8$, with $y_1, \dots, y_n \in [0, 1]$ fulfilling (19) for any $\tilde{\eta} > 0$, then, the distance matrix D^* is in $\mathcal{LDE}[(y_i), \sqrt{\tilde{\eta}}\alpha, \beta, 0, r]$ (as in Assumption L.1), for any $r \in (0, 1/4)$.*

Lemma L.3 states that, when F is a latent bi-Lipschitz matrix in $\mathcal{LBC}[(y_i), \alpha, \beta]$, the distance matrix D^* belongs to the class $\mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, 0, r]$ for the parameters $\tilde{\beta} = \beta$ and $\tilde{\alpha} = \alpha\sqrt{\tilde{\eta}}$.

Lemma L.4. *If F belongs to the class $\mathcal{LBC}[(y_i), \alpha, \beta]$ of latent bi-lipschitz matrices (as in Assumption 5.1), with $y_1, \dots, y_n \in [0, 1]$ fulfilling (19) for $\tilde{\eta} > 0$, then, Assumption L.2 holds for $\gamma = \tilde{\eta}\alpha$, any $r' \in (0, 1/4)$ and any $r'' \in [0, 1/8)$.*

The proofs of Lemma L.3 and L.4 are in appendix M.1.

Thus Theorem 5.1 (and L.1) follows from Theorem L.2. \square

In order to prove Theorem L.2, the rest of the section is organized as follows. We give the latent space extensions for the distance estimation in appendix L.3, for the first comparison matrix estimation in appendix L.4, for the second comparison matrix estimation in appendix L.5, for the final step (comparison matrix to permutation) in appendix L.6. Finally the proof of Theorem L.2 is in appendix L.7.

L.3 Extension for distance estimation

The next lemma upper bounds the (local) estimation errors of distances D_{ij}^* . This result is an extension of Lemma D.1 for general positions $y_i \in [0, 1]$. The proof is in appendix M.2.

Lemma L.5. *For any $8 \leq n$, and $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < \eta \leq r$, the following holds for any $y_1, \dots, y_n \in [0, 1]$ satisfying (18) and for any $D^* \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, 0, r]$ (as in Assumption L.1). With probability $1 - 1/n^4$, we have*

$$|\hat{D}_{ij} - D_{ij}^*| < \omega_{n,\eta} \quad \text{for all } i, j \in [n] \text{ such that } |y_{\pi_i} - y_{\pi_j}| \wedge D_{ij}^* \leq r \quad (123)$$

where

$$\omega_{n,\eta} = C \left(\sqrt{\tilde{\beta}\eta} + \sqrt{(\sigma+1)\sigma} \left(\frac{\log(n)}{n} \right)^{1/4} \right) \quad (124)$$

and C is a numerical constant.

In the special case where $y_i = i/n$, one has $\eta = 1/n$, and one recovers Lemma D.1. In contrast with Lemma D.1, the bound (124) depends on the spreading constant η of (18). In latent space models, the y_i 's are often assumed to be a uniform sample of $[0, 1]$, and hence the parameter η goes to zero (with n), and the error bound $\omega_{n,\eta}$ of (124) goes to zero. However, no assumption is made on the convergence of η in the current paper. Therefore, the estimator \widehat{D} may suffer from a bias $\sqrt{\tilde{\beta}\eta}$. When this bias is large, it may be interesting to change the estimator \widehat{D} for another estimator less sensitive to η . We propose such an alternative distance estimator in appendix O.

L.4 Extension for the first estimator of $H^{(\pi)}$

If the input D of `AgregLocalBisection` is in $\mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$, then the next proposition guarantees that the output H has an error smaller than a constant ρ . This result is a generalization of Proposition C.6 to the latent space setting.

Proposition L.6. *For any $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r$ and $0 \leq \omega$ and $0 < \eta$, the following holds for any $y_1, \dots, y_n \in [0, 1]$ as in (18), and for any $D \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$ (as defined in Assumption L.1) and any $\delta_1, \delta_2, \delta_3$ fulfilling (130). There exists $s \in \{\pm\}$ such that, the output H of `AgregLocalBisection` satisfies*

$$H_{ij} = sH_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } H_{ij} \neq 0 \text{ or } |y_{\pi_i} - y_{\pi_j}| \geq \rho ,$$

for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$

Compared to Proposition C.6, the constraints on $\delta_1, \delta_2, \delta_3$ change and now involve the spreading constant η . For the special case $\eta = 1/n$, one recovers Proposition C.6. The proof of Proposition L.6 is in appendix M.3.

L.5 Extension for the second estimator of $H^{(\pi)}$

Proposition L.7 extends Proposition C.7 to general $y_i \in [0, 1]$. There is no change between the two propositions. Indeed, going through the proof of Proposition C.7, we observe that none property of the y_i 's is used. As a consequence, the latent space formulation for general positions y_i follows directly.

Proposition L.7. *For any $0 < \gamma, r'$ and $0 \leq r''$ and $y_1 < \dots < y_n \in [0, 1]$, the following holds for any $F \in \mathcal{SCA}[(y_i), \gamma, r', r'']$ (as defined in Assumption L.2) and any $\tilde{\rho} \in [0, r' \wedge r'']$. If the the input H_1 of `LocalRefine` is deterministic or independent of the data A , with the following accuracy, for any $\epsilon \in \{\pm\}$,*

$$(H_1)_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } (H_1)_{ij} \neq 0 \text{ or } |y_{\pi_i} - y_{\pi_j}| \geq \tilde{\rho} , \quad (125)$$

then, with probability $1 - 4/n^3$, the output H of `LocalRefine` satisfies for all i, j ,

$$H_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{whenever } (H_1)_{ij} = 0 \text{ and } |y_{\pi_i} - y_{\pi_j}| \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}. \quad (126)$$

In addition, $H_{ij} = 0$ whenever $(H_1)_{ij} \neq 0$.

The next proposition is an extension of Proposition C.8.

Proposition L.8. *For any $0 < \tilde{\alpha} \leq \tilde{\beta}$ and $0 < r, \gamma, r'$ and $0 \leq \omega, r'', \tilde{\rho}$ and $0 < \eta$, the following holds for any $y_1 < \dots < y_n \in [0, 1]$ as in (18) and for any $D^* \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, 0, r]$ and $D \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$ (as defined in Assumption L.1), for any $F \in \mathcal{SCA}[(y_i), \gamma, r', r'']$ (as defined in Assumption L.2) and for any $\delta_1, \delta_2, \delta_3, \tilde{\rho}$ fulfilling (131). If the input H_1 of `LocalRefineWS` has the following accuracy, for any $\epsilon \in \{\pm\}$,*

$$(H_1)_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{for all } i, j \in [n] \text{ where } (H_1)_{ij} \neq 0 \text{ or } |y_{\pi_i} - y_{\pi_j}| \geq \tilde{\rho}, \quad (127)$$

then, with probability $1 - 8/n^2$, the output H satisfies

$$H_{ij} = \epsilon H_{ij}^{(\pi)} \quad \text{whenever } (H_1)_{ij} = 0 \text{ and } |y_{\pi_i} - y_{\pi_j}| \geq C \frac{\sigma}{\gamma} \sqrt{\frac{\log(n)}{n}}. \quad (128)$$

In addition, $H_{ij} = 0$ whenever $(H_1)_{ij} \neq 0$.

Compared to Proposition C.8, Proposition L.8 involves three changes: (1) the positions $x_i = y_{\pi_i}$'s in $[0, 1]$ are only assumed to satisfy (18); (2) Assumption 4.2 is replaced by Assumption L.2 w.r.t. the y_i 's; (3) the matrix D^*, D are \mathcal{LDE} w.r.t. the x_i 's as in Assumption L.1. The proof of Proposition L.8 is in appendix M.4.

L.6 Extension for final step (comparison matrix to permutation)

Generalizing the definition of error (33) for comparison matrix H , we say that H has an error smaller than ν , if it satisfies the following for an $s \in \{\pm\}$,

$$H_{ij} = s H_{ij}^{(\pi)} \quad \text{for all } i, j \text{ such that } |y_{\pi_i} - y_{\pi_j}| \geq \nu. \quad (129)$$

The next proposition generalizes Proposition C.9.

Proposition L.9. *Let $\nu > 0$. If a comparison matrix H has an error less than ν as in (129), then the permutation π^H in (38) has an L_{comp} error less than 2ν , that is $L_{\text{comp}}(\pi^H, \pi) \leq 2\nu$.*

The proof of Proposition L.9 is in appendix M.5.

L.7 Proof of Theorem L.2

The proof follows the same lines as for Theorem 4.1 (in appendix C.6). Since D^* is in $\mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, 0, r]$, we can apply Lemma L.5; this gives $\hat{D} \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega_{n,\eta}, r]$ where $\omega_{n,\eta}$ is defined in Lemma L.5. Then, Proposition L.6 and L.8 gives guarantees for $\hat{H}_1 + \hat{H}_2$ – the sum of the outputs of `AgregLocalBisection`(A, \hat{D}, \dots). Thus, the error of $\hat{H}_1 + \hat{H}_2$ is bounded by $(\sigma/\gamma)\sqrt{\log(n)/n}$, up to some numerical factor C . Finally, Proposition L.9 allows us to conclude that $\pi_\circ = \pi^{\hat{H}_1 + \hat{H}_2}$ satisfies the error bound of Theorem L.2.

L.8 Conditions on tuning parameters

The conditions on $\delta_1, \delta_2, \delta_3$ for the analysis of `AgregLocalBisection` in the latent space setting are written below. They are a direct extensions of the conditions (114) used for the matrix setting, where we have simply replaced n^{-1} by η .

$$\begin{aligned} \delta_1 &\geq \omega + \tilde{\beta}\eta, & \delta_2 &> \omega + \frac{\tilde{\beta}}{\tilde{\alpha}}(\delta_1 + \omega), & r &\geq \frac{\eta \vee (\delta_1 + \omega) \vee (\delta_2 + \omega)}{1 \wedge \tilde{\alpha}}, \\ \eta &\leq \rho \leq 1/8, & (r \wedge (\tilde{\alpha}/8)) - \omega &\geq \delta_3 > \omega + \tilde{\beta}\rho, \end{aligned} \quad (130)$$

for $\rho = (\delta_2 + \omega)/\tilde{\alpha}$.

We now give the conditions on $\delta_1, \delta_2, \delta_3$ for the analysis of `LocalRefineWS` in the latent space setting. They are extensions of the conditions (115), where we have included the modifications explained in the proof of Proposition L.8, in particular the replacement of ω_n with $\omega_{n,\eta}$.

Conditions (130) with the replacement $\omega \leftarrow (\omega \vee 2\omega_{n,\eta})$;

$$\tilde{\rho} \in [0, r'], \quad \tilde{\rho} \leq \rho, \quad \tilde{\rho}' + \rho \leq r'', \quad 2\eta + \tilde{\rho} + 2\tilde{\rho}' \leq 1/2 \quad 4\eta + 3\tilde{\rho} \leq 1; \quad (131)$$

$$\tilde{\omega} + \tilde{\beta}\rho < \delta_3, \quad \eta + \tilde{\rho} \leq r, \quad \phi_2 \leq \tilde{\beta}r, \quad (\phi_1 \vee \phi_2) + \tilde{\omega} \leq r;$$

for the notation $\tilde{\omega} = \tilde{\omega} \vee 2\omega_{n,\eta}$, and $\rho = (\delta_2 + \tilde{\omega})/\alpha$, and

$$\tilde{\rho}' = \frac{(\phi_1 \vee \phi_2) + 2\tilde{\omega}}{\tilde{\alpha}} \quad \text{with} \quad \phi_1 = \tilde{\beta}(\eta + \tilde{\rho}) \quad \text{and} \quad \phi_2 = \tilde{\beta}(\eta + \tilde{\alpha}^{-1}(\delta_3 + \tilde{\omega})).$$

M Proofs of lemmas and propositions from appendix L

M.1 Proofs of Lemma L.3 and L.4

We will use the notation $x_i = y_{\pi_i}$ for all i .

Proof of Lemma L.3. The proof follows the same lines as for Lemma C.1 (appendix D.1), up to the following difference. We replace $\frac{|\pi_i - \pi_j|}{n}$ by $|x_i - x_j| = |y_{\pi_i} - y_{\pi_j}|$ in the two main inequalities of the proof, and we lower bound the cardinal number $\#\{\ell : y_\ell \leq \frac{1}{4} + \frac{1}{n}\}$ by $\tilde{\eta}n$, using (19). \square

Proof of Lemma L.4. The proof follows the same lines as for Lemma C.2 (appendix D.1), up to the following difference. Since $y_j > y_i > \frac{3}{8}$, we have $y_i - r'' > \frac{1}{4}$ and then $\{\ell : y_\ell \leq \frac{1}{4}\} \subset \{\ell : y_\ell < y_i - r''\}$. This inclusion and condition (19) yield $\#\{\ell : y_\ell < y_i - r''\} \geq \tilde{\eta}n$. Combining with the assumption $F \in \mathcal{LBC}[(y_i), \alpha, \beta]$, we obtain

$$\sum_{\ell: y_\ell < y_i - r''} F_{i\ell} - F_{j\ell} \geq \alpha|y_i - y_j| \#\{\ell : y_\ell < y_i - r''\} \geq \alpha\tilde{\eta}n|y_i - y_j|.$$

Thus, Assumption L.2 holds for $\gamma = \alpha\tilde{\eta}$. \square

M.2 Proof of Lemma L.5

The proof is similar to that of Lemma D.1 in appendix E. The only difference is for controlling the bias term $D_{im_i}^*$ from the error bound in Lemma E.1. For $D^* \in \mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, 0, \sigma]$, we upper bound the bias term by $D_{im_i}^* \leq \tilde{\beta}\eta$ (instead of $D_{im_i}^* \leq \tilde{\beta}/n$).

M.3 Proof of Proposition L.6

Recall that, under Assumption 4.1, we used Lemma D.3 to D.5 for analysing `AgregLocalBisection` and proving guarantees for the output H (Proposition C.6). We now state the counterparts of these three lemmas for the general Assumption L.1.

Lemma M.1. *Let $y_1, \dots, y_n \in [0, 1]$ satisfy (18), and let $i \in [n]$. If the inputs in `LocalBisection`($i, D, \delta_1, \delta_2, \delta_3$) are such that, D is $\mathcal{LDE}[(y_i), \tilde{\alpha}, \tilde{\beta}, \omega, r]$ as in Assumption L.1, and $\delta_1, \delta_2, \delta_3$ fulfill the constraints (130)*

Then, the conclusion of Lemma D.3 holds.

Lemma M.2. *In the statement of Lemma D.4, replace the condition $n^{-1} \leq \rho$ by $\eta \leq \rho$, and “Lemma D.3” by “Lemma M.1”. Then, the same conclusion as in Lemma D.4 holds.*

Lemma M.3. *Under the hypotheses of Lemma M.2, the output H of `AgregLocalBisection` satisfies, for some $s \in \{\pm\}$,*

$$H_{ij} = sH_{ij}^{(\pi)} , \quad \text{for all } i, j \in [n] \text{ such that } H_{ij} \neq 0 \text{ or } |y_{\pi_i} - y_{\pi_j}| \geq \rho ,$$

where $\rho = (\delta_2 + \omega)/\tilde{\alpha}$.

Lemma M.1, M.2 and M.3 respectively generalize Lemma D.3, D.4 and D.5. The proofs of these three lemmas are in appendix N.

Proposition L.6 follows from these lemmas.

M.4 Proof of Proposition L.8

Reading the proof of Proposition C.8 (in appendix I), one can see the few spots where the y_i 's spreading comes into play; we will adjust them below. Ultimately, these minor changes will yield an extra constant (η) in the conditions on the tuning parameters $\delta_1, \delta_2, \delta_3$.

We will use the notation $x_i = y_{\pi_i}$ for all i . The equality $x_n = 1$ is replaced by the lower bound $x_n \geq 1 - \eta$, where η is the constant of the constraint (18). This replacement only occurs in the proof of Lemma I.1. Going through the proof of Lemma I.1, we observe that this change occurs in the following sub-Lemma I.5, I.4 I.2, I.3.

- To adapt Lemma I.5 for the latent space setting, we have to replace the constraint $\tilde{\rho} + 2t' \leq 1/2$ by $2\eta + \tilde{\rho} + 2t' \leq 1/2$. This change is useful at the end of proof of Lemma I.5, when proving the inequality $\tilde{D}_{kn} \geq \delta_3$. More precisely, we now have

$$x_n - x_k = x_n - x_i + (x_i - x_k) > (1 - \eta) - (1/2 + \tilde{\rho}/2) - t' = 1/2 - \eta - \tilde{\rho}/2 - t' \geq 1/4 ,$$

when $2\eta + \tilde{\rho} + 2t' \leq 1/2$.

- For Lemma I.4, the additive error $2\omega_n$ of the \mathcal{LDE} regularity is replaced by $2\omega_{n,\eta}$ where $\omega_{n,\eta}$ is defined in (124).
- For the generalization of Lemma I.2, we need to add the new constraint $4\eta + 3\tilde{\rho} \leq 1$. This is useful at the end of the proof of Lemma I.2. We now have

$$x_n - x_i \geq (1 - \eta) - (1/2 + \tilde{\rho}/2) \geq 1/4 ,$$

when $4\eta + 2\tilde{\rho} \leq 1$.

- For Lemma I.3, we replace n^{-1} by η . Thus, the constraint $n^{-1} + \tilde{\rho} \leq r$ becomes $\eta + \tilde{\rho} \leq r$, and the constants ϕ_1 and ϕ_2 are now equal to: $\phi_1 = \tilde{\beta}(\eta + \tilde{\rho})$ and $\phi_2 = \tilde{\beta}(\eta + \tilde{\alpha}^{-1}(\delta_3 + \omega))$.

Taking $t' = \tilde{\rho}'$ (as in the proof of Proposition C.8), we obtain the new conditions on $\delta_1, \delta_2, \delta_3$. They are gathered in (131), and used in Proposition L.8.

M.5 Proof of Proposition L.9

We will use the notation $x_i = y_{\pi_i}$ for all i . The proof follows the same lines as for Proposition C.9 (appendix D.5). We only have to replace the π_i/n 's with the x_i 's, and the L_{\max} loss by the L_{comp} loss. We just give the changes.

To obtain $L_{\text{comp}}(\pi^h, \pi) \leq 2\nu$ for π^h defined by (38), it is sufficient to prove that

$$\pi_j^h > \pi_i^h \quad \text{for all } i, j \in [n] \text{ such that } x_j \geq x_i + 2\nu .$$

Let $i, j \in [n]$ such that $x_j \geq x_i + 2\nu$. We introduce the following partition of the space $[0, 1]$,

$$\begin{aligned} I_1 &= [1, x_i - \nu], & I_2 &= (x_i - \nu, x_i + \nu), & I_3 &= [x_i + \nu, x_j - \nu], \\ I_4 &= (x_j - \nu, x_j + \nu), & I_5 &= [x_j + \nu, 1], \end{aligned}$$

assuming that $x_i > \nu$ and $x_j + \nu \leq 1$ (the other cases, where $x_i \leq \nu$ or $x_j + \nu > 1$, can be analyzed in a similar fashion, with a slight adaptation of the partition). We define the associated partition of indices $R_s = \{k \in [n] : x_k \in I_s\}$ for $s \in [5]$.

The rest of the proof can be done as for Proposition C.9 (appendix D.5). For completeness, we do the case $x_k \in I_2$. If $x_k \in I_2$, then $x_j - x_k \geq \nu$, and hence $H_{jk} = H_{jk}^{(\pi)} = 1$ for all $k \in R_2$. This yields

$$\sum_{k \in R_2} H_{jk} - H_{ik} = 1 + \sum_{k \in R_2, k \neq i} H_{jk} - H_{ik} \geq 1 ,$$

where we used $H_{ii} = 0$ in the equality, and $H_{ik} \in \{-1, 0, 1\}$ in the inequality.

N Proofs of Lemma M.1, M.2, M.3

Throughout the section, we use the notation $x_i = y_{\pi_i}$ for all i .

Proof of Lemma M.1. The proof follows the same lines as for Lemma D.3 (in appendix F.2), with a slight adaptation for the two following arguments: (1) the relation $|x_i - x_1| = |x_i - 1/n|$

is replaced by $|x_i - x_1| = |x_i - \eta|$, and similarly, $|x_i - x_n| = |1 - x_i|$ is replaced by $|x_i - x_n| = |(1 - \eta) - x_i|$; (2) the old $|x_\ell - x_{\ell_c}| \leq 1/n$ is replaced by $|x_\ell - x_{\ell_c}| \leq \eta$. The change (1) occurs in the last sentence of the proof of sub-Lemma F.6. The change (2) is in the last line of the proof of Lemma F.7.

Proof of Lemma M.2. The proof is almost the same as for Lemma D.4 (in appendix F.3), and we only point the changes out. Without the loss of generality, we assume that $\pi = id$, so that $x_1 < \dots < x_n$.

◦ *Point 2 of Align:* The definition of i_o changes: Let i_o such that $x_{i_o} \in [1/2 - \eta, 1/2]$. Such an i_o exists by (18). Then, following the same lines as in the old proof, we adapt the verification of the inequality $x_{i_o} - x_1 \geq 1/4$ as follows. Since $x_1 \in [0, \eta]$ by (18), we have $x_{i_o} - x_1 \geq (1/2 - \eta) - \eta \geq 1/2 - 2\eta \geq 1/4$ for $\eta \leq 1/8$.

◦ *Point 4 of Align:* The modifications concern the proof of (69) and (70), but we only present the changes for (69), the ones for (70) being similar. As a new definition of k_o , let $k_o \in [n]$ such that $x_{k_o} \in [3/4 - \eta, 3/4]$, whose existence is ensured by (18). To keep the same proof as before, we need to check two facts: (i) the inequality $x_{k_o} - x_i \geq \rho$ for $x_i < 1/4$, and (ii) $x_{k_o} \in [5/8, 3/4]$. For (i), we have $x_{k_o} - x_i \geq (3/4 - \eta) - 1/4 \geq 1/4 \geq \rho$, for $\eta \leq 1/4$ and $\rho \leq 1/8$. This (i) holds. For (ii), we have $x_{k_o} \in [(3/4 - \eta), 3/4] \subset [5/8, 3/4]$ for $\eta \leq 1/8$.

Instead of the point $x_{\lfloor n/2 \rfloor}$ used in the old proof, we define a new point x_ℓ satisfying $x_\ell \in [1/2 - \eta, 1/2]$. For this new point, we have to check two inequalities to keep the same old proof, namely, $x_\ell - x_i > 1/8$, and $x_{k_o} - x_\ell > 1/8$. Using $x_i < 1/4$, we indeed have

$$x_\ell - x_i > (1/2 - \eta) - 1/4 \geq 1/8 \quad \text{and} \quad x_{k_o} - x_\ell > 5/8 - 1/2 \geq 1/8 .$$

Proof of Lemma M.3. Same proof as for Lemma D.5 (in appendix F.4).

O Alternative distance estimator

Since the error bound on the estimator \widehat{D} has a bias term that is proportional to $\sqrt{\eta}$ (Lemma L.5), the estimator \widehat{D} may perform poorly, and so might the permutation estimator $\hat{\pi}_o$ build on \widehat{D} . This issue happens even if the noise level σ goes to zero. To rectify this, we describe in this section an alternative distance estimator whose estimation error goes to zero when $\sigma \rightarrow 0$.

Like in the definition of \widehat{D} , the crossed term in (22) is estimated by the scalar product $\langle A_i, A_j \rangle_n$. But, the quadratic term is now estimated by the empirical quadratic term $\langle A_i, A_i \rangle_n$. We have

$$\mathbb{E}[A_{ik}^2] \lesssim F_{\pi_i \pi_k}^2 + \sigma^2 .$$

It is therefore possible to use standard concentration bounds to prove that, with high probability, $|\langle A_i, A_i \rangle_n - \langle F_{\pi_i}, F_{\pi_i} \rangle_n| \lesssim \sigma^2 + c_n$ where

$$c_n = \left[\frac{|F|_{2,\infty}}{\sqrt{n}} + \sigma \right] \sigma \sqrt{\frac{\log n}{n}} . \quad (132)$$

It is the same upper bound as in Lemma E.2, up to an additive error σ^2 . Thus, when $\sigma \rightarrow 0$, the inner product $\langle A_i, A_i \rangle_n$ is a consistent estimator of the quadratic term $\langle F_{\pi_i}, F_{\pi_j} \rangle_n$.

Putting things together, we obtain the estimator $\widehat{D}_o^2(i, j) = \langle A_i, A_i \rangle_n + \langle A_j, A_j \rangle_n - 2\langle A_i, A_j \rangle_n$ of (22). Combining (132) and Lemma E.2, it is possible to prove that \widehat{D}_o^2 satisfies the following error bound

$$\max_{i, j \in [n]} \left| (D^*)_{ij} - (\widehat{D}_o^2)_{ij} \right| \lesssim \sigma^2 + \left[\frac{|F|_{2, \infty}}{\sqrt{n}} + \sigma \right] \sigma \sqrt{\frac{\log(n)}{n}} \quad (133)$$

with a probability greater than $1 - 1/n^4$.

Unlike the (main) estimator \widehat{D} , the alternative estimator \widehat{D}_o is consistent when $\sigma \rightarrow 0$, regardless of the value of η .

P Proof of Theorem 2.2

We recall that the lower bound is proved in the particular case where F is known and equal to the matrix F_0 ,

$$F_{0,ij} = 1 - \alpha \frac{|i - j|}{n}, \text{ for all } i, j \in [n]. \quad (134)$$

It is not difficult to check that, for $\alpha \in (0, 1]$ and any $\beta \geq \alpha$, the matrix F_0 belongs to $[0, 1]^{n \times n}$ as in model (2), and to the class $\mathcal{BL}[\alpha, \beta]$ of bi-Lipschitz matrices. We will establish the lower bound $(\sigma/\alpha)\sqrt{\log(n)/n}$ under the condition that $\alpha/\sigma \geq C_0\sqrt{\log(n)/n}$ where C_0 is a numerical constant (which will be set later). This last condition is satisfied as soon as $n \geq C_{\alpha, \sigma}$ for some constant $C_{\alpha, \sigma}$ only depending on α and σ .

Our minimax lower bound is based on Fano's method as stated below. We denote the set of permutations of $[n]$ by $\mathbf{\Pi}_n$. For two permutations π and π' in $\mathbf{\Pi}_n$, we denote the Kullback-Leibler divergence of $\mathbb{P}_{(F_0, \pi)}$ and $\mathbb{P}_{(F_0, \pi')}$ by $KL(\mathbb{P}_{(F_0, \pi)} \parallel \mathbb{P}_{(F_0, \pi')})$. Given the loss L_{\max} defined in (3), a radius $\epsilon > 0$ and a subset $\mathcal{S} \subset \mathbf{\Pi}_n$, the packing number $\mathcal{M}(\epsilon, \mathcal{S}, L_{\max})$ is defined as the largest number of points in \mathcal{S} that are at least ϵ away from each other with respect to L_{\max} . Below, we state a specific version of Fano's lemma.

Lemma P.1 (from [Yu, 1997]). *Consider any subset $\mathcal{S} \subset \mathbf{\Pi}_n$. Define the Kullback-Leibler diameter of \mathcal{S} by*

$$d_{KL}(\mathcal{S}) = \sup_{\pi, \pi' \in \mathcal{S}} KL(\mathbb{P}_{(F_0, \pi)} \parallel \mathbb{P}_{(F_0, \pi')}) .$$

Then, for any estimator $\hat{\pi}$ and any $\epsilon > 0$, we have

$$\sup_{\pi \in \mathcal{S}} \mathbb{P}_{(F_0, \pi)} \left[L_{\max}(\hat{\pi}, \pi) \geq \frac{\epsilon}{2} \right] \geq 1 - \frac{d_{KL}(\mathcal{S}) + \log(2)}{\log \mathcal{M}(\epsilon, \mathcal{S}, L_{\max})} .$$

In view of the above proposition, we mainly have to choose a suitable subset \mathcal{S} , control its Kullback-Leibler diameter, and get a sharp lower bound of its packing number. A difficulty stems from the fact that the loss $L_{\max}(\hat{\pi}, \pi)$ is invariant when reversing the ordering π .

Let $k := C_1(\sigma/\alpha)\sqrt{n \log(n)}$, for a small enough numerical constant $C_1 \in (0, 1]$ (which will be set later). To ensure that $k \leq n/4$, we enforce the condition $\alpha/\sigma \geq C_0\sqrt{\log(n)/n}$, with $C_0 := 4C_1$. For simplicity, we assume that $n/4$ is an integer (otherwise take the floor value

$\lfloor n/4 \rfloor$). We introduce $n/4$ permutations $\pi^{(s)} \in \mathbf{\Pi}_n$, $s = 1, \dots, n/4$. For each $s \in [n/4]$, let $\pi_j^{(s)}$ be such that

$$\forall j \in [n] \setminus \{s, s+k\} : \pi_j^{(s)} = j, \quad \text{and} \quad \pi_s^{(s)} = s+k, \quad \text{and} \quad \pi_{s+k}^{(s)} = s .$$

Each permutation $\pi^{(s)}$ is therefore equal to the identity $(j)_{j \in [n]}$ up to an exchange of the two indices s and $s+k$. This collection of $n/4$ permutations is denoted by

$$\mathcal{S} := \{\pi^{(1)}, \dots, \pi^{(n/4)}\} . \quad (135)$$

For the subset $\mathcal{S} \subset \mathbf{\Pi}_n$, we readily check that

$$\forall s, t \in \left[\frac{n}{4} \right], s \neq t : \quad L_{\max}(\pi^{(t)}, \pi^{(s)}) \geq \frac{k}{n} .$$

This gives a lower bound on the packing number $\mathcal{M}(\epsilon_n, \mathcal{S}, L_{\max})$ of radius ϵ_n :

$$\mathcal{M}(\epsilon_n, \mathcal{S}, L_{\max}) \geq n/4 , \quad \text{for } \epsilon_n = k/n .$$

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 P.2. *For any $\pi, \pi' \in \mathbf{\Pi}_n$, and any $n \times n$ matrix F , we have $KL(\mathbb{P}_{(F,\pi)} \parallel \mathbb{P}_{(F,\pi')}) \leq \frac{1}{2\sigma^2} \sum_{i,j \in [n]} (F_{\pi_i \pi_j} - F_{\pi'_i \pi'_j})^2$.*

Combining with the definition (134) of F_0 , we obtain

$$KL(\mathbb{P}_{(F_0,\pi)} \parallel \mathbb{P}_{(F_0,\pi')}) \leq C_2 n \frac{(\alpha \epsilon_n)^2}{\sigma^2} = C_2 C_1^2 \log(n) ,$$

for $\epsilon_n = k/n = C_1(\sigma/\alpha)\sqrt{\log(n)/n}$, and a numerical constant $C_2 > 0$. Taking the value $C_1 = (2\sqrt{C_2})^{-1}$, we have

$$d_{KL}(\mathcal{S}) \leq \frac{\log(n)}{4} .$$

Applying Lemma P.1 to the set \mathcal{S} of (135), we arrive at

$$\inf_{\hat{\pi}} \sup_{\pi \in \mathcal{S}} \mathbb{P}_{(F_0,\pi)} \left[L_{\max}(\hat{\pi}, \pi) \geq \frac{\epsilon_n}{2} \right] \geq 1 - \frac{\log(n)/4 + \log(2)}{\log(n/4)} \geq \frac{1}{2} ,$$

as soon as n is greater than some numerical constant. The lower bound $\epsilon_n/2$ is of the order of $(\sigma/\alpha)\sqrt{\log(n)/n}$. Theorem 2.2 follows. \square

P.0.1 Proof of Claim P.2

We recall that $\mathbb{P}_{(F,\pi)}$ is the probability distribution of the data A . For all pairs $i, j \in [n]$, we denote the marginal distribution of A_{ij} by $\mathbb{P}_{(F_{ij},\pi)}$. By definition of the Kullback-Leibler divergence, we have

$$KL(\mathbb{P}_{(F,\pi)} \parallel \mathbb{P}_{(F,\pi')}) = \sum_{i < j} KL(\mathbb{P}_{(F_{ij},\pi)}, \mathbb{P}_{(F_{ij},\pi')}) \leq \sum_{i,j} \frac{(F_{\pi_i \pi_j} - F_{\pi'_i \pi'_j})^2}{2\sigma^2} ,$$

where the equality follows immediately from the independence of the A_{ij} , $i < j$, and the inequality from Claim P.3 (for $d = 1$, $\Sigma_1 = \Sigma_2 = \sigma^2$, $\mu_1 = F_{ij}$, $\mu_2 = F'_{ij}$).

Claim P.2 is proved. \square

Claim P.3. *Let two (multivariate) normal distributions $P = N(\mu_1, \Sigma_1)$ and $Q = N(\mu_2, \Sigma_2)$, both d dimensional, with respective mean vectors $\mu_1, \mu_2 \in \mathbb{R}^d$, and covariance matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$. Then, we have*

$$KL(P, Q) = \frac{1}{2} \left[\log \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right) + (\mu_1 - \mu_2) \Sigma_2^{-1} (\mu_1 - \mu_2) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - d \right] .$$

In particular, when $\Sigma_1 = \Sigma_2 = \sigma \text{Id}$, we have the simpler formula

$$KL(P, Q) = \frac{\|\mu_1 - \mu_2\|_2^2}{2\sigma^2} .$$

We used the notation Id for the identity matrix, $|M|$ for the determinant and $\text{tr}(M)$ for the trace of any matrix M .

For a proof of the first relation in Claim P.3, see [Duchi, 2014, section 9]. The second relation is a direct consequence.

References

- [Atkins et al., 1998] 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(1):297–310.
- [Barnard et al., 1995] Barnard, S. T., Pothen, A., and Simon, H. (1995). A spectral algorithm for envelope reduction of sparse matrices. Numerical linear algebra with applications, 2(4):317–334.
- [Braverman and Mossel, 2009] Braverman, M. and Mossel, E. (2009). Sorting from noisy information. arXiv preprint arXiv:0910.1191.
- [Briend et al., 2024] Briend, S., Giraud, C., Lugosi, G., and Sulem, D. (2024). Estimating the history of a random recursive tree. arXiv preprint arXiv:2403.09755.
- [Cai and Ma, 2022] Cai, T. T. and Ma, R. (2022). Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. arXiv preprint arXiv:2201.06438.
- [Chatterjee and Mukherjee, 2019] Chatterjee, S. and Mukherjee, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. IEEE Transactions on Information Theory, 65(6):3525–3539.
- [Chen et al., 2019] Chen, Y., Fan, J., Ma, C., and Wang, K. (2019). Spectral method and regularized mle are both optimal for top-k ranking. Annals of statistics, 47(4):2204.

- [Collier and Dalalyan, 2016] Collier, O. and Dalalyan, A. S. (2016). Minimax rates in permutation estimation for feature matching. The Journal of Machine Learning Research, 17(1):162–192.
- [De Castro et al., 2017] De Castro, Y., Lacour, C., and Ngoc, T. M. P. (2017). Adaptive estimation of nonparametric geometric graphs. arXiv preprint arXiv:1708.02107.
- [Diaconis and Janson, 2007] Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs. arXiv preprint arXiv:0712.2749.
- [Diaz et al., 2020] Diaz, J., McDiarmid, C., and Mitsche, D. (2020). Learning random points from geometric graphs or orderings. Random Structures & Algorithms, 57(2):339–370.
- [Duchi, 2014] Duchi, J. (2014). Derivations for linear algebra and optimization. University of California.
- [Elson et al., 2004] Elson, J., Karp, R. M., Papadimitriou, C. H., and Shenker, S. (2004). Global synchronization in sensor networks. In Farach-Colton, M., editor, LATIN 2004: Theoretical Informatics, pages 609–624, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Flammarion et al., 2019] Flammarion, N., Mao, C., and Rigollet, P. (2019). Optimal rates of statistical seriation. Bernoulli, 25(1):623–653.
- [Fogel et al., 2014] Fogel, F., d’Aspremont, A., and Vojnovic, M. (2014). Serialrank: Spectral ranking using seriation. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, Advances in Neural Information Processing Systems, volume 27. Curran Associates, Inc.
- [Fogel et al., 2013] Fogel, F., Jenatton, R., Bach, F., and d’Aspremont, A. (2013). Convex relaxations for permutation problems. In Advances in Neural Information Processing Systems, pages 1016–1024.
- [Fulkerson and Gross, 1965] Fulkerson, D. and Gross, O. (1965). Incidence matrices and interval graphs. Pacific journal of mathematics, 15(3):835–855.
- [Galstyan et al., 2022] Galstyan, T., Minasyan, A., and Dalalyan, A. S. (2022). Optimal detection of the feature matching map in presence of noise and outliers. Electronic Journal of Statistics, 16(2):5720–5750.
- [Garriga et al., 2011] Garriga, G. C., Junttila, E., and Mannila, H. (2011). Banded structure in binary matrices. Knowledge and information systems, 28(1):197–226.
- [Giraud et al., 2021] Giraud, C., Issartel, Y., and Verzelen, N. (2021). Localization in 1d nonparametric latent space models from pairwise affinities. arXiv preprint arXiv:2108.03098.
- [Giridhar and Kumar, 2006] 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, pages 4915–4920.
- [Hoff et al., 2002] 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(460):1090–1098.

- [Issartel, 2021] Issartel, Y. (2021). On the estimation of network complexity: Dimension of graphons. Journal of Machine Learning Research, 22(191):1–62.
- [Janssen and Smith, 2020] Janssen, J. and Smith, A. (2020). Reconstruction of line-embeddings of graphons. arXiv preprint arXiv:2007.06444.
- [Lovász, 2012] Lovász, L. (2012). Large networks and graph limits, volume 60. American Mathematical Soc.
- [Mao et al., 2018a] Mao, C., Pananjady, A., and Wainwright, M. J. (2018a). Towards optimal estimation of bivariate isotonic matrices with unknown permutations. arXiv preprint arXiv:1806.09544.
- [Mao et al., 2018b] Mao, C., Weed, J., and Rigollet, P. (2018b). Minimax rates and efficient algorithms for noisy sorting. In Algorithmic Learning Theory, pages 821–847. PMLR.
- [Natik and Smith, 2021] Natik, A. and Smith, A. (2021). Consistency of spectral seriation. arXiv preprint arXiv:2112.04408.
- [Penrose, 2003] Penrose, M. (2003). Random geometric graphs, volume 5. Oxford university press.
- [Recanati et al., 2017] Recanati, A., Bröls, T., and d’Aspremont, A. (2017). A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. Bioinformatics, 33(20):3188–3194.
- [Recanati et al., 2018] Recanati, A., Kerdreux, T., and d’Aspremont, A. (2018). Reconstructing latent orderings by spectral clustering. arXiv preprint arXiv:1807.07122.
- [Rigollet and Hütter, 2023] Rigollet, P. and Hütter, J.-C. (2023). High-dimensional statistics. arXiv preprint arXiv:2310.19244.
- [Robinson, 1951] Robinson, W. S. (1951). A method for chronologically ordering archaeological deposits. American Antiquity, 16(4):293–301.
- [Rocha et al., 2018] Rocha, I., Janssen, J., and Kalyaniwalla, N. (2018). Recovering the structure of random linear graphs. Linear Algebra and its Applications, 557:234–264.
- [Yu, 1997] Yu, B. (1997). Assouad, Fano, and le Cam. In Festschrift for Lucien Le Cam, pages 423–435. Springer.