Active Seriation: Efficient Ordering Recovery with Statistical Guarantees

James Cheshire **Yann Issartel**

LTCI, Télécom Paris, Institut Polytechnique de Paris

Abstract

We consider the problem of active seriation, where the goal is to recover an unknown ordering of n items based on noisy observations of pairwise similarities. The similarities are assumed to correlate with the underlying ordering: pairs of items that are close in the ordering tend to have higher similarity scores, and vice versa. In the active setting, the learner sequentially selects which item pairs to query and receives noisy similarity measurements. We propose a novel active seriation algorithm that provably recovers the correct ordering with high probability. Furthermore, we provide optimal performance guarantees in terms of both the probability of error and the number of observations required for successful recovery.

1 Introduction

The seriation problem involves ordering n items based on noisy measurements of pairwise similarities. This reordering problem originates in archaeology, where it was used for the chronological dating of graves [Robinson, 1951]. More recently, it has found applications in data science across various domains, including envelope reduction for sparse matrices [Barnard et al., 1995], read 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], and interval graph identification [Fulkerson and Gross, 1965]. In many of these settings, pairwise measurements can be made in an active or adaptive manner, leveraging information from previously chosen pairs. Motivated by these applications, we study the problem of recovering an accurate item ordering from a sequence of actively selected pairwise measurements.

In the seriation paradigm, we assume the existence of an unknown symmetric matrix M representing pairwise similarities between a collection of n items. This similarity matrix is structured so that the similarities are correlated with an unknown underlying ordering of the items. This ordering can be encoded by a permutation $\pi = (\pi_1, \dots, \pi_n)$ of $[n]$, where the similarity M_{ij} between items i and j tends to be large when their positions π_i and π_j are close, and small when they are far apart.

To model this structure formally, the literature often assumes that M is a permuted Robinson matrix (also known as an R-matrix) [Fogel et al., 2013, Recanati et al., 2018, Janssen and Smith, 2020, Giraud et al., 2021]. A matrix is called Robinson if its rows and columns are unimodal and attain their maxima along the main diagonal. Under this assumption, the similarity M is modeled as a Robinson matrix whose rows and columns have been permuted according to π.

In the active seriation problem, observations are collected sequentially. At each step, a pair of items (i, j) is selected, and a noisy observation of the corresponding similarity M_{ij} is obtained. This observation is modeled as a realization of a σ-sub-Gaussian random variable with mean M_{ij}. A total of T pairs are selected, where T is fixed in advance, and the choice of each pair may depend on the outcomes of previous selections. It is standard to assume that the T observations are mutually independent.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

The goal is to recover the hidden ordering π of the n items from these noisy, actively chosen observations. To evaluate the performance of an estimator $\hat{\pi}$, we consider the probability of error $\mathbb{P}\{\hat{\pi} \neq \pi\}$, which measures the likelihood of failing to recover the true ordering.

Our first contribution is a characterization of the optimal rates for the active seriation problem on the class of similarity matrices whose pairwise similarities are separated by a minimal gap Δ. The statistical difficulty of ordering recovery is captured by the quantity

$$\text{SNR} := \frac{\Delta^2 T}{\sigma^2 n} , \tag{1}$$

which can be interpreted as the number of observations per item, T/n , multiplied by the signal-to-noise ratio (SNR) per observation, Δ^2/σ^2. Our theorems identify two regimes depending on the value of this SNR. For conciseness, we focus on the challenging regime where the SNR per observation satisfies $\frac{\Delta^2}{\sigma^2} \leq 1$, and use the notation $\lesssim$ and $\gtrsim$ to ignore absolute constant factors.

- **The Impossible Regime:** SNR $\lesssim \ln n$. In this regime, we prove that no algorithm can recover the ordering with a vanishing probability of error.
- **The Recovery Regime:** SNR $\gtrsim \ln n$. We show that the optimal recovery rate is achievable in polynomial time, with high probability. Specifically, we show that the probability of error decreases exponentially fast with the SNR, and no algorithm can achieve a faster decay, establishing optimality.

Our second contribution is algorithmic. We introduce *Active Seriation by Iterative Insertion* (ASII), an active procedure for estimating the ordering π. Unlike most existing seriation methods, which are non-active, ASII is remarkably simple, runs in polynomial time, yet achieves optimality guarantees both in terms of the probability of error and the number of observations required for recovery.

Our third contribution extends our results to general Robinson matrices, without any Δ-separation assumption. We introduce a robust variant of our procedure that adaptively focuses on well-separated items, and show that it can still recover a consistent partial ordering with high confidence.

Finally, we illustrate the performance of ASII through numerical experiments and a real-data application. On synthetic data, ASII remains competitive with specialized batch algorithms on Toeplitz matrices and clearly outperforms them on more general non-Toeplitz structures. On single-cell RNA sequencing data, ASII produces biologically coherent reorderings, highlighting the robustness and practical relevance of the proposed approach beyond idealized settings.

1.1 Related work

Classical seriation. The classical (non-active) seriation problem was first addressed by [Atkins et al., 1998] in the noiseless setting, using a spectral algorithm. Later works analyzed this algorithm under noise [Fogel et al., 2013, Giraud et al., 2021, Natik and Smith, 2021], showing good performance under strong structural assumptions [Giraud et al., 2021]. More recent contributions proposed polynomial-time algorithms with statistical guarantees under various assumptions, such as Lipschitz structure [Giraud et al., 2021], relaxed average Lipschitz conditions [Issartel et al., 2024], Toeplitz structure [Cai and Ma, 2022, Berenfeld et al., 2024], and others [Janssen and Smith, 2020]. In contrast, our analysis does not rely on additional structural assumptions such as Lipschitz or Toeplitz conditions.

Statistical-computational gaps. Prior studies have suggested statistical–computational gaps in the non-active seriation problem [Giraud et al., 2021, Cai and Ma, 2022, Berenfeld et al., 2024], where known polynomial-time algorithms fall short of achieving statistically optimal rates under specific noise conditions or structural assumptions. While some of these gaps have recently been closed [Issartel et al., 2024], the resulting algorithms tend to be complex and may not scale well in practice. In the active setting, however, we show that a simple and computationally efficient algorithm achieves statistically optimal performance.

Active and adaptive observation models. While classical seriation assumes access to all pairwise similarities, our work considers an active setting where similarities are observed sequentially, based on past observations. Related problems include adaptive ranking and sorting under noisy comparisons, such as active preference learning [Jamieson and Nowak, 2011], adaptive sorting [Braverman and

Mossel, 2009], and ranking from pairwise comparisons [Heckel et al., 2019]. However, these problems rely on pairwise comparisons (e.g., is item i preferred to item j?) to infer a total order of items. In contrast, seriation builds on pairwise similarity scores that encode proximity in an unknown underlying ordering. This distinction leads to different statistical and algorithmic challenges.

Connections to bandit models. In classical multi-armed bandits (MAB) [Bubeck and Cesa-Bianchi, 2012], each arm yields independent rewards, and the goal is to maximize cumulative reward or identify the best arm. In contrast, in active seriation, each query (i, j) measures the similarity between two interdependent items, and all measurements must be consistent with a single latent ordering. This interdependence prevents a direct application of standard MAB algorithms such as UCB or Thompson Sampling, which treat arms as independent and do not exploit structural relationships between them.

Algorithmically, our approach is related to noisy binary search and thresholding bandits [Feige et al., 1994, Karp and Kleinberg, 2007, Ben-Or and Hassidim, 2008, Nowak, 2009, Emamjomeh-Zadeh et al., 2016, Cheshire et al., 2020, Cheshire et al., 2021], which rely on adaptive querying under uncertainty. However, these methods operate on low-dimensional parametric models, whereas seriation involves a combinatorial ordering that must remain globally consistent across item pairs.

Relation to active ranking problems. A related but distinct problem is active ranking [Heckel et al., 2019], where a learner infers a total order based on noisy pairwise comparisons or latent score estimates. Extensions include Borda, expert, and bipartite ranking [Saad et al., 2023, Cheshire et al., 2023]. These methods typically assume that each item is associated with an intrinsic scalar score, and that pairwise feedback expresses a directional preference between items. In contrast, seriation relies on symmetric pairwise similarity information, which encodes proximity rather than preference. Recovering the latent ordering therefore requires global consistency among all pairwise similarities, making the problem more constrained and structurally different from standard ranking tasks.

2 Problem formulation

In this section, we formally state the seriation problem considered in this paper.

2.1 Pairwise similarities, Robinson structure, and orderings

Given a collection of items $[n] := \{1, \dots, n\}$, let $M = [M_{ij}]_{1 \le i,j \le n}$ denote their (unknown) symmetric *similarity matrix*, where the coefficient $M_{ij} \in \mathbb{R}$ measures the similarity between items i and j. Our structural assumption on M is related to the class of *Robinson matrices*, introduced below. The entries of a Robinson matrix decrease as one moves away from the (main) diagonal; see Figure 1.

Definition 2.1. *A matrix $R \in \mathbb{R}^{n \times n}$ is called a Robinson matrix (or R-matrix) if it is symmetric and satisfies the inequalities*

$$R_{i+1,j} > R_{i,j} \quad \text{and} \quad R_{i,j} > R_{i,j+1}$$

for all $1 \le i < j \le n$ on the upper triangle.

Thus, each row and column of an R-matrix is unimodal and attains its maximum on the diagonal. Following [Atkins et al., 1998], a matrix is said to be *pre-R* if it can be transformed into an R-matrix by simultaneously permuting its rows and columns (see Figure 1). In this paper, we assume the similarity matrix M is pre-R, i.e.,

$$M = R_\pi := [R_{\pi_i, \pi_j}]_{1 \le i,j \le n} \tag{2}$$

for some R-matrix R and some permutation $\pi = (\pi_1, \dots, \pi_n)$ of $[n]$. We call π an *ordering* of the n items with similarity matrix M if it satisfies (2) for some R-matrix.

In (2), the similarities M_{ij} reflect the ordering π in that M_{ij} tends to be larger when the positions π_i and π_j are close together, and smaller when they are far apart.

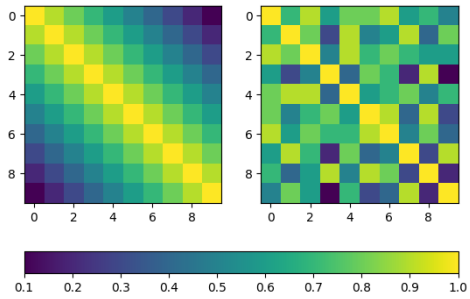

Figure 1: R-matrix & a permuted version.

Remark 1. *The items have two orderings: if π is an ordering, then the reverse permutation π^{rev}, defined by $\pi_i^{\mathrm{rev}} = n + 1 - \pi_i$ for all i, is also an ordering. Indeed, one can verify that $M = R^{\mathrm{rev}}_{\pi^{\mathrm{rev}}}$ where R^{rev} is the R-matrix defined by $R^{\mathrm{rev}}_{ij} := R_{n+1-i,n+1-j}$.*

A key quantity in our analysis is the separation between consecutive entries in the underlying R-matrix. Specifically, for any pre-R matrix M as in (2), we define the *minimal gap* as

$$\Delta_M := \min_{1 \le i < j \le n-1} \{(R_{i+1,j} - R_{i,j}) \wedge (R_{i,j} - R_{i,j+1})\} \quad , \tag{3}$$

where we write $a \wedge b := \min(a, b)$. This quantity measures the smallest difference between neighboring entries in the R-matrix associated with M. (Note that Δ_M is well-defined, even though M admits two associated R-matrices: R and its reverse R^{rev}.)

For future reference, define

$$\mathcal{M}_\Delta := \left\{ M \in \mathbb{R}^{n \times n} : M \text{ is pre-R, and } \Delta_M \ge \Delta \right\} \tag{4}$$

as the set of pre-R matrices with minimal gap at least Δ.

To simplify the presentation of our findings, we focus on the challenging regime where

$$\frac{\Delta}{\sigma} \le 1, \tag{5}$$

i.e., the signal-to-noise ratio per observation is at most 1. This excludes mildly stochastic regimes where the problem has essentially the same difficulty as in the noiseless case.

2.2 The active seriation problem

In the active seriation problem, the algorithm sequentially queries T pairs of items and receives noisy observations of their pairwise similarities. The unknown similarity matrix $M \in \mathcal{M}_\Delta$ (see (4)) encodes these similarities, and the goal is to recover an ordering associated with M as in (2).

At each round $t = 1, \ldots, T$, the algorithm selects a pair (i_t, j_t) with $i_t \neq j_t$, possibly depending on the outcomes of previous queries. It then receives a noisy observation of the similarity $M_{i_t j_t}$, modeled as a realization of a σ-sub-Gaussian random variable[1] with mean $M_{i_t j_t}$. This sub-Gaussian framework covers several standard observation models, including Gaussian noise, Bernoulli observations, and bounded variables.

After T queries, the algorithm outputs a permutation $\hat{\pi}$ of $[n]$ as its estimate of the true ordering. The algorithm is considered successful if $\hat{\pi}$ recovers either π or its reverse π^{rev}, as both induce valid Robinson structures on the similarity matrix (see Remark 1). The probability of error is thus defined as:

$$p_{M,T} \;:=\; \mathbb{P}_{M,T} \left\{ \hat{\pi} \neq \pi \text{ and } \hat{\pi} \neq \pi^{\mathrm{rev}} \right\}, \tag{6}$$

where the probability is over the randomness in the T observations collected on M.

3 Seriation procedure

ASII (Active Seriation by Iterative Insertion) reconstructs the underlying ordering by iteratively inserting each new item into an already ordered list. At iteration k, given a current ordering $\pi^{(k-1)}$ of $k-1$ items, the algorithm inserts item k at its correct position to form an updated ordering $\pi^{(k)}$. This process is repeated until the full ordering of all n items is obtained.

To perform this insertion efficiently, two key subroutines are used:

(i) **Local comparison rule.** To decide where to insert k, the algorithm must compare its position relative to items already in the list. This is achieved through the subroutine TEST, which determines whether k lies to the left, in the middle, or to the right of two reference items.

[1] We recall that a real-valued random variable X is said to be *σ-sub-Gaussian* if $\mathbb{E}[\exp(uX)] \le \exp\left(\frac{u^2\sigma^2}{2}\right)$ for all $u \in \mathbb{R}$.

(ii) **Efficient insertion strategy.** To minimize the number of comparisons, the algorithm performs a binary search over the current ordering, where each comparison is made via TEST. Because these tests are noisy, the procedure is further stabilized through a backtracking mechanism, which allows the algorithm to detect and correct occasional local errors without increasing the overall sample complexity.

The next paragraphs describe these two subroutines in more detail.

Local comparison rule. Given three items (k, l, r) such that $\pi_l < \pi_r$, the goal is to determine whether π_k lies to the left, in the middle, or to the right of (π_l, π_r) in the unknown ordering:

$$\pi_k < \pi_l \qquad \textbf{vs} \qquad \pi_k \in (\pi_l, \pi_r) \qquad \textbf{vs} \qquad \pi_k > \pi_r \ . \tag{7}$$

The intuition follows from a basic property of Robinson matrices: when k lies between l and r, its similarity to both l and r tends to be higher than the similarity between l and r themselves. Hence, by comparing the three empirical similarities $\widehat{M}_{kl}$, $\widehat{M}_{kr}$, and $\widehat{M}_{lr}$, we identify the smallest one as corresponding to the pair of items that are farthest apart. Formally, if $\widehat{M}_{lr}$ is the smallest, k lies in the middle; if $\widehat{M}_{kr}$ is the smallest, k lies to the left of (l, r); and if $\widehat{M}_{kl}$ is the smallest, k lies to the right. The subroutine TEST returns $b \in \{-1, 0, 1\}$ accordingly.

In practice, TEST uses a small fixed number of samples T_0 to make each test sampling-efficient, except in a few critical steps of ASII (e.g., initialization of a binary search) where higher reliability is required. This sampling-efficient design trades individual test accuracy for global robustness: occasional misclassifications are later corrected by the backtracking mechanism. The formal pseudocode of TEST is provided in Appendix B.

Efficient insertion strategy. The idea of incorporating backtracking into a noisy binary search has appeared in several algorithmic studies, e.g., [Feige et al., 1994, Ben-Or and Hassidim, 2008, Emamjomeh-Zadeh et al., 2016], and was further explored in [Cheshire et al., 2021]. Here, we adopt this general principle to design a robust insertion mechanism that remains reliable under noisy relational feedback.

To efficiently insert a new element into an already ordered list, we use the subroutine BINARY & BACKTRACKING SEARCH (BBS), which determines the relative position of item k within the current ordering $(\pi_1, \ldots, \pi_{k-1})$. The search proceeds by repeatedly testing whether k lies in the left or right half of a candidate interval, thereby narrowing down the possible insertion range. Because test outcomes are noisy, even a single incorrect decision can misguide the search and lead to an erroneous final placement. A naive fix would be to allocate many samples per test to ensure highly reliable outcomes, but this would increase the sample complexity and undermine the benefit of active sampling. Instead, BBS uses a small, constant number of samples per test—enough to ensure a constant success probability (e.g., around $3/4$). This design trades individual test accuracy for global sample efficiency, with backtracking acting as a corrective mechanism that restores robustness.

The algorithm keeps track of previously explored intervals and performs sanity checks at each step to detect inconsistencies in the search path. When an inconsistency is detected, it backtracks to an earlier interval and resumes the search. This prevents local mistakes from propagating irreversibly. Theoretical analysis shows that as long as the number of correct local decisions outweighs the number of incorrect ones—an event that occurs with high probability— the final insertion position is accurate. Hence, the backtracking mechanism enables the algorithm to combine low-sample local testing with global reliability, providing both efficiency and robustness despite noisy observations. A pseudocode of this procedure is provided in Appendix B.

4 Results for Δ-separated matrices

We analyze the fundamental limits of ordering recovery over the class $\mathcal{M}_\Delta$, deriving both information-theoretic lower bounds and algorithmic upper bounds on the error probability. These results are expressed in terms of the signal-to-noise ratio (SNR), defined earlier in (1), and identify two distinct regimes separated by the threshold $\log n$.

4.1 Impossibility regime

When the SNR satisfies $\frac{\Delta^2 T}{\sigma^2 n} \lesssim \ln n$, no algorithm can recover the ordering with vanishing error probability. The following theorem formalizes this impossibility, establishing a constant lower bound on the error probability (6) for any algorithm in this regime.

Theorem 4.1 (Impossibility regime)**.** *Let $n \geq 9$, $\sigma > 0$, $\Delta > 0$, and $T \geq 1$. If the SNR satisfies*

$$\frac{\Delta^2 T}{\sigma^2 n} \leq c_0 \ln n,$$

for a sufficiently small absolute constant $c_0 > 0$, then for any algorithm, there exists a matrix $M \in \mathcal{M}_\Delta$ such that

$$p_{M,T} \geq \frac{1}{2}.$$

Intuitively, the impossibility regime is more pronounced when the minimal gap Δ is small or when the sub-Gaussian parameter σ is large. The proof is provided in Appendix G.

4.2 Recovery regime

In the complementary regime, where the SNR satisfies $\frac{\Delta^2 T}{\sigma^2 n} \gtrsim \ln n$, we establish performance guarantees for ASII. Specifically, part (a) of the next theorem provides an upper bound on the error probability (6) of ASII, which decreases exponentially fast with the SNR. Conversely, part (b) provides a matching lower bound, showing that no algorithm can achieve a faster error decay than exponential in the SNR.

Theorem 4.2 (Recovery regime)**.** *Let $n \geq 4$, $\sigma > 0$, $\Delta > 0$, and $T \geq 1$. If the SNR satisfies*

$$\frac{\Delta^2 T}{\sigma^2 n} \geq c_1 \ln n,$$

for a sufficiently large absolute constant $c_1 > 0$, and condition (5) *holds, then:*

(a) For any $M \in \mathcal{M}_\Delta$, the error probability of ASII *satisfies*

$$p_{M,T} \leq \exp\left(-\tfrac{1}{800}\,\frac{\Delta^2 T}{\sigma^2 n}\right). \tag{8a}$$

(b) For any seriation algorithm, there exists $M \in \mathcal{M}_\Delta$ such that

$$p_{M,T} \geq \exp\left(-8\,\frac{\Delta^2 T}{\sigma^2 n}\right). \tag{8b}$$

Theorem 4.2 completes the characterization of the recovery regime, showing that the error probability of ASII decays exponentially once the SNR $\frac{\Delta^2 T}{\sigma^2 n}$ exceeds a logarithmic threshold in n. Together, Theorems 4.1 and 4.2 delineate the statistical landscape of active seriation, establishing a sharp phase transition between impossibility and recovery at the critical SNR level $\frac{\Delta^2 T}{\sigma^2 n} \asymp \ln n$. Proofs of parts (a) and (b) are given in Appendices E and G, respectively.

Instance-dependent guarantee. For simplicity, Theorem 4.2 (a) is stated for a fixed Δ, but the same bound holds instance-wise in terms of the true minimal gap Δ_M of M. Specifically, for any $M \in \mathcal{M}_0$ (the class of pre-R matrices), if $\ln n \lesssim \frac{\Delta_M^2 T}{\sigma^2 n}$, then

$$p_{M,T} \;\leq\; \exp\left(-\frac{\Delta_M^2 T}{800\sigma^2 n}\right).$$

Hence, the performance of the algorithm automatically scales with each instance, without requiring any knowledge of Δ_M, σ, or the SNR.

4.3 Optimal sample complexity for exact recovery

The following corollary summarizes Theorems 4.1 and 4.2 in terms of sample complexity. Combining the impossibility result of Theorem 4.1 with the recovery guarantee of Theorem 4.2, we obtain a precise characterization (up to absolute constants) of the number of observations required for exact recovery.

Corollary 4.3 (Optimal sample complexity)**.** *The minimax-optimal number of observations required for exact recovery with probability at least* $1 - 1/n^2$ *satisfies*

$$T^\star \asymp \frac{\sigma^2 n \ln n}{\Delta^2}.$$

This result follows directly from Theorem 4.2 (a), which guarantees exponentially small error probability whenever $T \gtrsim \frac{\sigma^2 n \ln n}{\Delta^2}$, and from Theorem 4.1, which shows that recovery is impossible below this threshold. Hence, ASII attains the minimax-optimal sample complexity $T^\star$, which depends transparently on the problem parameters (Δ, σ, n). This expression formalizes the intuition that recovery becomes easier as the signal gap Δ increases or the noise parameter σ decreases. Importantly, it also shows that exact recovery is achievable with a number of queries $T \ll n^2$, significantly improving over the classical batch setting where all pairwise similarities must be observed.

4.4 Intrinsic hardness and invariance to model assumptions

Both lower bounds (Theorems 4.1 and 4.2 (b)) are established under a Gaussian noise model with centered, homoscedastic entries of variance σ^2, whereas our upper bound is proved in the more general sub-Gaussian setting allowing heterogeneous noise levels. Since these bounds match up to constant factors, potential heterogeneity in the noise variances does not affect the minimax rates.

Moreover, the lower bounds are derived for the simple Toeplitz family $R_{ij} = \Delta(n - |i - j|)$, yet the attainable rates coincide with those obtained under the general assumption $M \in \mathcal{M}_\Delta$. Hence, allowing heterogeneous, non-Toeplitz matrices and sub-Gaussian noise comes at no statistical cost.

Finally, even in the extreme case where the Robinson matrix R is fully known —such as the one-parameter family $R_{ij} = \Delta(n - |i - j|)$— the attainable rates remain unchanged. This shows that the hardness of active seriation stems from the combinatorial nature of the latent ordering rather than from uncertainty on the similarity model.

5 Extension to general Robinson matrices

In many practical scenarios, some items may be nearly indistinguishable in terms of their pairwise similarities. For instance, a few rows (or columns) of the similarity matrix may be almost identical. In such cases, the minimal gap Δ_M can become extremely small. When this occurs, some items are effectively indistinguishable in the presence of noise, making their exact ordering statistically impossible to recover. To handle this situation, we now consider arbitrary Robinson matrices $M \in \mathcal{M}_0$, without any separation assumption. Our goal is then to recover the ordering of a subset of items that are *sufficiently well separated*, while ignoring indistinguishable ones.

5.1 Algorithmic adaptation: simplified seriation procedure

The extended algorithm, denoted EXT-ASII(Δ), follows the same iterative insertion framework as ASII, but employs a simplified binary search and a more conservative local testing rule to handle potentially non–Δ-separated items. Backtracking is no longer required, as each local decision is made with high confidence by allocating $\ln n$ times as many samples per test as in ASII.

Modified test rule. Given a triplet of items (ℓ, r, k) and a tolerance parameter $\Delta > 0$, the subroutine TEST_Δ determines whether k lies to the left, in the middle, or to the right of (ℓ, r) based on their empirical similarities. Let $\widehat{M}_{\ell r}$, $\widehat{M}_{\ell k}$, and $\widehat{M}_{rk}$ denote the corresponding sample means, each computed from $T_0 = \lfloor T/(3\,n\,(\log_2(n) + 1)) \rfloor$ observations per pair. Define the event

$$I[(a, b), x, \Delta] \;:=\; \big\{\, \widehat{M}_{ab} + \tfrac{\Delta}{2} < \widehat{M}_{xa} \wedge \widehat{M}_{xb} \,\big\}.$$

The output $b \in \{-1, 0, 1\}$ of TEST_Δ is assigned as follows:

$$b = \begin{cases} 0 & \text{if } I[(\ell, r), k, \Delta] \quad \text{(middle)}, \\ -1 & \text{if } I[(k, r), \ell, \Delta] \quad \text{(left)}, \\ +1 & \text{if } I[(k, \ell), r, \Delta] \quad \text{(right)}. \end{cases} \tag{9}$$

If none of the three cases above applies, the item k is considered Δ–close to at least one of ℓ or r and is discarded from further insertion.

Binary search without backtracking. At iteration k, the algorithm considers item k and attempts to insert it into the current estimated ordering $\pi^{(k-1)}$ over the active subset $S^{(k-1)} \subset [k-1]$ of items successfully inserted so far. The insertion relies on a standard binary search guided by TEST_Δ. If a test fails to decide —that is, if k is Δ-close to a boundary element— the item is discarded and not further inserted. Otherwise, the binary search proceeds within $\pi^{(k-1)}$ until the insertion position of k is determined.

Output. After n iterations, the algorithm outputs a total ordering $\hat{\pi}$ over the subset $S = S^{(n)}$ of successfully inserted items. Items rejected during the procedure are those empirically Δ-close to others, thereby ensuring that S forms a Δ-maximal subset as defined in Section 5.2.

This simplified variant preserves the structure of ASII while adapting test confidence to weaker separation, yielding the guarantees of Theorem 5.2.

5.2 Recovery guarantees

Here $\Delta > 0$ denotes a user-chosen tolerance parameter, which sets the resolution at which the algorithm attempts to distinguish items. The guarantees below hold without any prior knowledge of the true matrix M or the noise level σ.

Formally, given $\Delta > 0$, we define the notion of a Δ-*maximal subset*. For a subset $S \subset [n]$ and a matrix $M \in \mathbb{R}^{n \times n}$, we write M_S for the submatrix of M consisting of entries M_{ij} such that $i, j \in S$.

Definition 5.1 (Δ-maximal subset)**.** *A subset $S \subset [n]$ is said to be Δ-maximal for a matrix $M \in \mathcal{M}_0$ if it cannot be enlarged by adding any single item without violating the Δ-separation condition, that is,*

$$M_{S \cup \{k\}} \notin \mathcal{M}_\Delta \quad \textit{for all } k \in [n] \setminus S.$$

Intuitively, S cannot be enlarged without adding items too similar (within Δ) to those already in S.

Given such a subset S, we aim to recover the induced ordering π_S of the true permutation π restricted to S, or its reverse π_S^{rev}. Let $p_{M,T,S}$ denote the probability of misordering S after T active queries:

$$p_{M,T,S} \;:=\; \mathbb{P}_{M,T}\{\hat{\pi} \neq \pi_S \text{ and } \hat{\pi} \neq \pi_S^{\text{rev}}\}\,. \tag{10}$$

Theorem 5.2 (Extension to any Robinson matrix)**.** *Let $n \geq 1$, $\sigma > 0$, $\Delta > 0$, and $T \geq 1$. If*

$$\frac{\Delta^2 T}{\sigma^2 n} \;\geq\; c_1 (\ln n)^2,$$

for a sufficiently large absolute constant $c_1 > 0$, there exists an absolute constant $c_2 > 0$ such that the following holds.

For any $M \in \mathcal{M}_0$, the output $\hat{\pi}$ of the extended algorithm EXT-ASII*(Δ) is a permutation over a subset $S \subset [n]$ satisfying:*

- *S is Δ-maximal in the sense of Definition 5.1;*
- *the error probability obeys the bound*

$$p_{M,T,S} \;\leq\; \exp\!\left(-\,c_2 \frac{\Delta^2 T}{\sigma^2 n \ln n}\right). \tag{11}$$

Compared to Theorem 4.2 (a), the error exponent in (11) includes an additional logarithmic factor. This factor stems from using $\ln n$ times as many samples per test for high-confidence decisions. A refined analysis, possibly reusing the backtracking technique of Theorem 4.2 (a), should remove this factor and yield optimal rates; we leave this refinement for future work. Overall, Theorem 5.2 shows that reliable partial recovery remains achievable even without uniform Δ-separation.

6 Empirical illustration

We illustrate the behavior of ASII through numerical experiments and a real-data example.

Numerical simulations. We assess the empirical performance of ASII on synthetic data and compare it to three benchmark methods: (i) the batch seriation algorithm ADAPTIVE SAMPLING [Cai and Ma, 2022], (ii) the SPECTRAL SERIATION algorithm [Atkins et al., 1998], and (iii) a naive insertion variant without backtracking. All methods are evaluated under identical sampling budgets on four synthetic scenarios, including both Toeplitz and non-Toeplitz Robinson matrices (see Appendix C for full experimental details and figures).

On Toeplitz matrices, ASII performs slightly below specialized batch methods such as ADAPTIVE SAMPLING and SPECTRAL SERIATION, which are tailored to that structure, yet still achieves high empirical accuracy. In contrast, on non-Toeplitz matrices, ASII consistently outperforms both ADAPTIVE SAMPLING and SPECTRAL SERIATION, which may fail entirely in some settings. These experiments confirm that ASII maintains reliable performance beyond classical structural assumptions, demonstrating strong robustness across heterogeneous matrix geometries.

Application to real data. We further assess the robustness of ASII on real single-cell RNA sequencing data (human primordial germ cells, from [Guo et al., 2015], previously analyzed by [Cai and Ma, 2022]). Although such biological data depart substantially from the idealized Robinson models assumed in our theory, ASII still produces a meaningful reordering of the empirical similarity matrix, revealing coherent developmental trajectories among cells. This illustrates the practical relevance and robustness of the proposed approach beyond the assumptions made in our theoretical model. Full experimental details are provided in Appendix C.

7 Discussion

Scope and noise regimes. Our analysis focused on the stochastic regime where the per-observation signal-to-noise ratio Δ/σ is at most 1, which captures the most challenging setting for active seriation. The results, however, extend naturally to less noisy regimes ($\Delta/\sigma > 1$): in that case, accurate recovery requires only $T \gtrsim n \log n$ queries, reflecting the intrinsic $O(n \log n)$ cost of performing n adaptive binary insertions. Further details and discussion are provided in Appendix A.1.

Gain from active learning. Our active framework enables recovery of the underlying ordering without observing the entire similarity matrix. Whereas batch approaches such as spectral methods require $O(n^2)$ observations, our active algorithm ASII succeeds with only $T \gtrsim (\sigma^2/\Delta^2) n \log n$ samples. This corresponds to a fraction $(\sigma^2/\Delta^2)(\log n/n)$ of the full matrix and yields a substantial reduction in sample complexity, provided that the signal-to-noise ratio Δ^2/σ^2 remains bounded away from zero. This gain arises from the ability of adaptive sampling to draw information from a well-chosen, small subset of pairwise similarities, from which the entire matrix can be reordered, achieving strong statistical efficiency under limited sampling budgets.

Potential applications. Beyond synthetic settings, seriation techniques are broadly relevant in domains where pairwise similarity information reflects a latent one-dimensional structure. Examples include genomic sequence alignment, where seriation helps reorder genetic fragments by similarity, and recommendation systems, where item-item similarity matrices can reveal latent preference orderings. We also illustrated, on real single-cell RNA sequencing data (human primordial germ cells from [Guo et al., 2015]), that ASII can recover biologically meaningful trajectories despite the data departing strongly from our theoretical model. These settings often involve noisy or costly pairwise measurements, making active seriation particularly appealing as a statistically efficient alternative to batch reordering methods.

Comparison with batch methods. In certain scenarios —especially when differences between candidate orderings are highly localized— our active formulation can succeed under weaker signal conditions than batch methods. This contrast reflects a design difference: batch approaches often rely on global matrix discrepancies, which can impose stronger separation requirements in such localized settings, whereas our active algorithm concentrates sampling on locally ambiguous regions. A detailed comparative example is provided in Appendix A.2.

Fixed-budget and fixed-confidence formulations. Throughout this work, we focused on the *fixed-budget* setting, where the total number of samples T is fixed in advance and the objective is to minimize the error probability within this budget. The algorithm does not require prior knowledge of the noise variance σ nor of the minimal signal gap Δ; it simply allocates the available budget T across tests. As a result, the performance depends on σ and Δ only through the signal-to-noise ratio $\mathrm{SNR} = \Delta^2 T/(\sigma^2 n)$, which determines the achievable accuracy.

A natural extension would be to study the complementary *fixed-confidence* setting, where the algorithm must adaptively decide when to stop sampling in order to achieve a prescribed confidence level. Such a formulation would typically require variance-aware sampling policies and data-driven stopping rules, possibly involving online estimation of σ. Developing such an adaptive, fixed-confidence version of ASII remains an interesting direction for future work.

Conclusion. This work introduces the first active-learning formulation of the seriation problem, together with sharp theoretical guarantees and a simple, polynomial-time algorithm. We characterize a phase transition in sample complexity governed by the signal-to-noise ratio $\mathrm{SNR} := \Delta^2 T/(\sigma^2 n)$: recovery is impossible when $\mathrm{SNR} \lesssim \log n$, while ASII achieves near-minimax optimality once this threshold is exceeded. Our analysis highlights how adaptive sampling and corrective backtracking improve statistical efficiency.

Beyond Δ-separated instances, we further extended the analysis to arbitrary Robinson matrices, introducing a user-chosen tolerance parameter Δ that sets the resolution at which recovery is attempted. At this scale, ASII reliably recovers the ordering of all items that are sufficiently well separated, formalized through the notion of Δ-maximal subsets. Nevertheless, some matrices may exhibit global ordering structure even in the absence of any locally separated subsets. Capturing such globally structured signals —beyond the local, Δ-based framework— remains an open and important direction for future work.

Acknowledgements

The work of J. Cheshire is supported by the FMJH, ANR-22-EXES-0013.

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.

[Ben-Or, 1983] Ben-Or, M. (1983). Lower bounds for algebraic computation trees. In Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing, STOC '83.

[Ben-Or and Hassidim, 2008] Ben-Or, M. and Hassidim, A. (2008). The bayesian learner is optimal for noisy binary search (and pretty good for quantum as well). In 2008 49th Annual IEEE Symposium on Foundations of Computer Science, pages 221–230. IEEE.

[Berenfeld et al., 2024] Berenfeld, C., Carpentier, A., and Verzelen, N. (2024). Seriation of t {\oe} plitz and latent position matrices: optimal rates and computational trade-offs. arXiv preprint arXiv:2408.10004.

[Braverman and Mossel, 2009] Braverman, M. and Mossel, E. (2009). Sorting from noisy information. arXiv preprint arXiv:0910.1191.

[Bubeck and Cesa-Bianchi, 2012] Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armedbandit problems.

[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.

[Cheshire and Clémençon, 2025] Cheshire, J. and Clémençon, S. (2025). Active bipartite ranking with smooth posterior distributions. In The 28th International Conference on Artificial Intelligence and Statistics.

[Cheshire et al., 2023] Cheshire, J., Laurent, V., and Clémençon, S. (2023). Active bipartite ranking. In Thirty-seventh Conference on Neural Information Processing Systems.

[Cheshire et al., 2020] Cheshire, J., Ménard, P., and Carpentier, A. (2020). The influence of shape constraints on the thresholding bandit problem. arXiv preprint arXiv:2006.10006.

[Cheshire et al., 2021] Cheshire, J., Ménard, P., and Carpentier, A. (2021). Problem dependent view on structured thresholding bandit problems. In International Conference on Machine Learning, pages 1846–1854. PMLR.

[Cormen et al., 2009] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). Introduction to Algorithms, Third Edition. The MIT Press, 3rd edition.

[Elson et al., 2004] Elson, J., Karp, R. M., Papadimitriou, C. H., and Shenker, S. (2004). Global synchronization in sensornets. In Farach-Colton, M., editor, LATIN 2004: Theoretical Informatics, pages 609–624, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Emamjomeh-Zadeh et al., 2016] Emamjomeh-Zadeh, E., Kempe, D., and Singhal, V. (2016). Deterministic and probabilistic binary search in graphs. In Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, pages 519–532. ACM.

[Falahatgar et al., 2017] Falahatgar, M., Orlitsky, A., Pichapati, V., and Suresh, A. T. (2017). Maximum selection and ranking under noisy comparisons. In Proceedings of the 34th International Conference on Machine Learning.

[Feige et al., 1994] Feige, U., Raghavan, P., Peleg, D., and Upfal, E. (1994). Computing with noisy information. SIAM Journal on Computing, 23(5):1001–1018.

[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.

[Garivier et al., 2019] Garivier, A., Ménard, P., and Stoltz, G. (2019). Explore first, exploit next: The true shape of regret in bandit problems. Mathematics of Operations Research, 44(2):377–399.

[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.

[Gerchinovitz et al., 2020] Gerchinovitz, S., Ménard, P., and Stoltz, G. (2020). Fano's inequality for random variables. Statistical Science, 35(2):178–201.

[Giraud et al., 2021] Giraud, C., Issartel, Y., and Verzelen, N. (2021). Localization in 1d non-parametric 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.

[Gu and Xu, 2023] Gu, Y. and Xu, Y. (2023). Optimal bounds for noisy sorting. In Proceedings of the 55th Annual ACM Symposium on Theory of Computing, pages 1502–1515.

[Guo et al., 2015] Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and dna methylome landscapes of human primordial germ cells. Cell, 161(6):1437–1452.

[Hao et al., 2023] Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology.

[Heckel et al., 2019] Heckel, R., Shah, N. B., Ramchandran, K., and Wainwright, M. J. (2019). Active ranking from pairwise comparisons and when parametric assumptions do not help.

[Issartel et al., 2024] Issartel, Y., Giraud, C., and Verzelen, N. (2024). Minimax optimal seriation in polynomial time. arXiv preprint arXiv:2405.08747.

[Jamieson and Nowak, 2011] Jamieson, K. G. and Nowak, R. (2011). Active ranking using pairwise comparisons. Advances in neural information processing systems, 24.

[Janssen and Smith, 2020] Janssen, J. and Smith, A. (2020). Reconstruction of line-embeddings of graphons. arXiv preprint arXiv:2007.06444.

[Karp and Kleinberg, 2007] Karp, R. M. and Kleinberg, R. (2007). Noisy binary search and its applications. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 881–890. Society for Industrial and Applied Mathematics.

[Lattimore and Szepesvári, 2020] Lattimore, T. and Szepesvári, C. (2020). Bandit algorithms. Cambridge University Press.

[Natik and Smith, 2021] Natik, A. and Smith, A. (2021). Consistency of spectral seriation. arXiv preprint arXiv:2112.04408.

[Nowak, 2009] Nowak, R. (2009). The geometry of generalized binary search. arXic preprint arXiv:0910.4397.

[Recanati et al., 2017] Recanati, A., Brüls, 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.

[Ren et al., 2019] Ren, W., Liu, J. K., and Shroff, N. (2019). On sample complexity upper and lower bounds for exact ranking from noisy comparisons. Advances in Neural Information Processing Systems, 32.

[Robinson, 1951] Robinson, W. S. (1951). A method for chronologically ordering archaeological deposits. American Antiquity, 16(4):293–301.

[Saad et al., 2023] Saad, E. M., Verzelen, N., and Carpentier, A. (2023). Active ranking of experts based on their performances in many tasks. In Proceedings of the 40th International Conference on Machine Learning.

[Soch et al., 2023] Soch et al., J. (2023). The book of statistical proofs.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: All theoretical results referenced in abstract are clearly stated in main text.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Yes both theoretical and practical limitations are discussed

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Assumptions are clearly stated in main text with corresponding proofs in appendix

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All code used to generate empirical results is provided.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: All code used to produce empirical results is provided

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: ´Experimental section is clearly explained. All code used to produce empirical results is provided

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error bars are provided.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Experiments are purely illustrative as this is a theoretically focused work

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: No ethical concerns

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Limited potential impact as the work is theoretically focused

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: No risk posed

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Competing algorithms clearly referenced

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: No crowdsourcing or research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: NA

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: No involvement of LLMs

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.