On the Ranking Recovery from Noisy Observations up to a Distortion

Minoh Jeong⋆, Martina Cardone⋆, Alex Dytso†
⋆ University of Minnesota, Minneapolis, MN 55455, USA, Email: {jeong316, mcardone}@umn.edu
† New Jersey Institute of Technology, Newark, NJ 07102, USA, Email: alex.dytso@njit.edu

***Abstract*—This paper considers the problem of recovering the ranking of a data vector from noisy observations, up to a distortion. Specifically, the noisy observations consist of the original data vector corrupted by isotropic additive Gaussian noise, and the distortion is measured in terms of a distance function between the estimated ranking and the true ranking of the original data vector. First, it is shown that an optimal (in terms of error probability) decision rule for the estimation task simply outputs the ranking of the noisy observation. Then, the error probability incurred by such a decision rule is characterized in the low-noise regime, and shown to grow sublinearly with the noise standard deviation. This result highlights that the proposed *approximate* version of the ranking recovery problem is significantly less noise-dominated than the *exact* recovery considered in [Jeong, ISIT 2021].**

I. Introduction

Today, ranking data is a pervading task in several applications, such as search engines [1], biomedical [2], recommender systems [3], feature matching [4], and communication systems [5]. However, the data might be *noisy*, e.g., because of privacy considerations [6], users might desire to privatize it with the addition of some noise, before sharing it with an external data collector. Thus, it is paramount to understand the impact of the noise on the performance of the ranking task.

In this paper, we propose an *approximate* version of the ranking recovery problem that we introduced and studied in [4], [7]. In particular, the problem in [4], [7] consists of recovering the *exact* permutation (also ranking) according to which an input data vector was sorted before being corrupted by some additive noise. Here, we study a relaxed version of this problem, namely we allow for some controlled distortion in the estimation of the ranking (see Section II). In particular, we focus on the case where the noise is isotropic Gaussian, and we measure the distortion in terms of a distance function between the estimated ranking and the true ranking of the original data vector. We first show (see Section III) that an optimal (in terms of error probability) decision rule for this problem is given by the *linear decoder* proposed in [4], which consists of simply declaring the ranking of the noisy observation. This decoder has a low complexity, namely polynomial in the dimension of the data. Then (see Section IV), we study the probability of error incurred in the low-noise regime when the linear decoder is used. In particular, we show that the error probability grows sublinearly with the noise standard deviation σ. This is a notable difference with respect to the exact version of the ranking recovery problem in [7], where we showed that the error probability grows linearly with σ. This result highlights that the proposed approximate ranking recovery problem is significantly less noise-dominated with respect to exact recovery. All our derived results hold under mild assumptions on the distance function and are satisfied by widely used distance functions, such as the Hamming distance and the Kendall's tau rank distance.

The work of M. Jeong and M. Cardone was supported in part by the U.S. National Science Foundation under Grant CCF-1849757.

A. Related Work

Recently, permutation relevant estimation problems have gained significant importance and are investigated in a large body of the literature [8]–[23]. For a joint Gaussian distribution, the ranking estimation problem was studied in [8]–[11]. A pairwise ordering estimation for the bivariate case was considered in [8], and its extended version to an arbitrary dimension n was studied in [9]. The results were generalized from the Gaussian assumption to an elliptically contoured distribution in [10], [11]. In particular, the authors characterized the condition for the covariance matrix that maximizes the probability of correctness of such estimation problems using the minimum mean square error (MMSE) estimator. Besides, the MMSE estimator under the Gaussian noise assumption was shown to be the only linear estimator that achieves the minimum probability of error for the exact permutation (also ranking) recovery problem in [4]. The estimation of a sorted vector from some noisy observations was proposed in [23]. In particular, the authors showed that the MMSE estimator on the sorted data can be decomposed as a linear combination of estimators on the unsorted data.

II. Notation and Problem Formulation

A. Notation and Definitions

Boldface upper case letters $\mathbf{X}$ denote vector/matrix random variables; the boldface lower case letter $\mathbf{x}$ indicates a specific realization of $\mathbf{X}$; $[n_1 : n_2]$ is the set of integers from n_1 to $n_2 \geq n_1$; I_n is the identity matrix of dimension n; $\mathbf{0}_n$ is the column vector of dimension n of all zeros. Calligraphic letters indicate sets; $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$; $\mathbb{1}_{\mathcal{A}}$ is the indicator function; $\mathcal{P}_n$ is the set of all permutations of an n-dimensional vector; $\stackrel{d}{=}$ denotes equality in distribution; $f_X(x)$ is the probability density function of a random variable X.

Definition 1. *We denote by* $\pi_{\mathbf{x}}$ *the permutation of* $\mathbf{x} \in \mathbb{R}^n$ *such that*

$$(\mathbf{x}_{\pi_{\mathbf{x}}})_1 \leq ... \leq (\mathbf{x}_{\pi_{\mathbf{x}}})_i \leq ... \leq (\mathbf{x}_{\pi_{\mathbf{x}}})_n,$$

where $\mathbf{x}_\tau$ *is the sorted version of* $\mathbf{x}$ *according to* $\tau \in \mathcal{P}_n$*, and* $(\mathbf{x}_\tau)_i$ *is the* i*-th element of* $\mathbf{x}_\tau$*, with* $i \in [1:n]$*.*

We denote by $\mathbf{r_x}$ *the ranking*[1] *of* $\mathbf{x} \in \mathbb{R}^n$ *such that* $(\mathbf{r_x})_i$ *indicates that* x_i *is the* $(\mathbf{r_x})_i$*-th smallest among the entries of* $\mathbf{x}$*. The set of all rankings of size* n *is denoted by* $\mathcal{R}_n$*.*

Example 1. *If* $\mathbf{x} = (-2, 3, -6, 1, 2)$*, then we have*

$$\pi_{\mathbf{x}} = (3, 1, 4, 5, 2), \text{ and } \mathbf{r_x} = (2, 5, 1, 3, 4). \qquad (1)$$

Definition 2. *For any two rankings* $\mathbf{r_u} \in \mathcal{R}_n$ *and* $\mathbf{r_v} \in \mathcal{R}_n$*, the Hamming distance between* $\mathbf{r_u}$ *and* $\mathbf{r_v}$ *is defined as*

$$\mathsf{d}_H(\mathbf{r_u}, \mathbf{r_v}) = |\{i : (\mathbf{r_u})_i \neq (\mathbf{r_v})_i\}| = \sum_{i=1}^{n} \mathbb{1}_{\{(\mathbf{r_u})_i \neq (\mathbf{r_v})_i\}}.$$

Definition 3. *For any two rankings* $\mathbf{r_u} \in \mathcal{R}_n$ *and* $\mathbf{r_v} \in \mathcal{R}_n$*, the Kendall's tau rank distance between* $\mathbf{r_u}$ *and* $\mathbf{r_v}$ *is defined as*

$$\mathsf{d}_K(\mathbf{r_u}, \mathbf{r_v}) = |\{(i, j) : i < j, \\ \operatorname{sgn}((\mathbf{r_u})_i - (\mathbf{r_u})_j) \neq \operatorname{sgn}((\mathbf{r_v})_i - (\mathbf{r_v})_j)\}|,$$

where sgn *denotes the sign function.*

Example 2. *Let* $n = 4$ *and* $\mathbf{r_x} = (1, 3, 2, 4)$*. To have* $\mathsf{d}_H(\mathbf{r_x}, \mathbf{r_y}) = 2$ *we need*

$$\mathbf{r_y} \in \{(1, 3, 4, 2), (1, 4, 2, 3), (1, 2, 3, 4), \\ (3, 1, 2, 4), (2, 3, 1, 4), (4, 3, 2, 1)\},$$

whereas to have $\mathsf{d}_K(\mathbf{r_x}, \mathbf{r_y}) = 1$*, we need*

$$\mathbf{r_y} \in \{(2, 3, 1, 4), (1, 2, 3, 4), (1, 4, 2, 3)\}.$$

B. Preliminaries and Known Results

We consider the following model,

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \qquad (2)$$

where $\mathbf{X} \in \mathbb{R}^n$ is any exchangeable random vector[2] and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, with $\mathbf{X}$ and $\mathbf{N}$ being independent.

In this work, we study the ranking recovery problem where the goal is to estimate $\mathbf{r_x}$ given the noisy observation $\mathbf{y}$ under the model in (2). In [4], [7], we considered this problem under an *exact* ranking recovery constraint[3], i.e., we were interested in recovering the *exact* ranking according to which the input data vector $\mathbf{X}$ was sorted. In particular, in [4], [7] we showed that the decision rule, referred to as decoder in the remaining of the paper, that minimizes the error probability consists of declaring $\mathbf{r_x}$ to be equal to the ranking $\mathbf{r_y}$ of $\mathbf{y}$. Because of this structure, this optimal (in terms of error probability) decision rule was referred to as *linear decoder*. More formally, let $\phi : \mathbb{R}^n \to \mathcal{R}_n$ denote the decoder. In [4], [7], we showed that, for any value of the noise standard deviation σ, for the *exact* ranking recovery we have that

$$\phi_{\text{lin}} \in \operatorname*{argmin}_{\phi} P_e(\phi, \sigma) = \operatorname*{argmin}_{\phi} \Pr(\phi(\mathbf{Y}) \neq \mathbf{r_X}), \qquad (3)$$

where $\phi_{\text{lin}}(\mathbf{y}) = \mathbf{r_y}$, and $P_e(\phi, \sigma) = \Pr(\phi(\mathbf{Y}) \neq \mathbf{r_X})$ is the probability of error incurred for $\sigma \in \mathbb{R}_+$ when the decoder ϕ is applied. In [7], we also characterized $P_e(\phi_{\text{lin}}, \sigma)$ in the low-noise (i.e., $\sigma \to 0$), and high-noise (i.e., $\sigma \to \infty$) regimes. Notably, in the low-noise regime, we showed that the probability of error is linear in σ, i.e., $P_e(\phi_{\text{lin}}, \sigma) \approx c\sigma$ with a coefficient c that can be proportional to n^2. This result shows that the exact ranking recovery problem is noise dominated and hence, the estimation task can be difficult to implement in practice. Followed by this observation, a natural question arises: *How does the approximate recovery problem, where a fixed number of errors are allowed, perform?* We next formally define the *approximate* ranking recovery problem.

C. Approximate Ranking Recovery

Different from the exact ranking recovery, in the *approximate* version of the problem, a fixed number of errors is allowed in the recovery of the ranking $\mathbf{r_x}$. To formulate this problem, we let $\mathsf{d} : \mathcal{R}_n^2 \to \mathbb{R}_+$ be a distance function, which measures the distance (e.g., Hamming in Definition 2, Kendall's tau in Definition 3) between two rankings. In particular, in order to consider a proper distance function, d has to satisfy the following two assumptions:

A1: $\mathsf{d}(\mathbf{r_u}, \mathbf{r_v}) = 0$ if and only if $\mathbf{r_u} = \mathbf{r_v}$; and
A2: $\mathsf{d}(\mathbf{r_u}, \mathbf{r_v}) = \mathsf{d}(P\mathbf{r_u}, P\mathbf{r_v})$ for any permutation matrix P.

We then define the ball with respect to d centered at $\mathbf{r_c} \in \mathcal{R}_n$ with radius ℓ, namely

$$\mathcal{B}_{\mathsf{d}}(\mathbf{r_c}, \ell) = \{\mathbf{r_x} \in \mathcal{R}_n : \mathsf{d}(\mathbf{r_c}, \mathbf{r_x}) \leq \ell\}, \qquad (4)$$

where ℓ is referred to as distortion threshold and denotes the maximum number of errors that are allowed.

Example 3. *Consider the case* $n = 4$*, for which* $|\mathcal{R}_4| = 24$*. Let* d *be the Hamming distance in Definition 2, and* $\ell = 2$*. For* $\mathbf{r_c} = (1, 2, 3, 4)$*, we have that*

$$\mathcal{B}_{\mathsf{d}_H}(\mathbf{r_c}, 2) = \{(1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (2, 1, 3, 4), \\ (3, 2, 1, 4), (4, 2, 3, 1), (1, 4, 3, 2)\},$$

where the first ranking in $\mathcal{B}_{\mathsf{d}_H}(\mathbf{r_c}, 2)$ *is* $\mathbf{r_c}$ *(hence, it has zero Hamming distance), whereas all the other rankings are at Hamming distance equal to two from* $\mathbf{r_c}$*.*

The *approximate* ranking recovery problem is the estimation task for which we seek to recover $\mathbf{r_x}$ with a certain distortion, measured by d, up to a threshold equal to ℓ. This problem can also be seen as estimating $\hat{\mathbf{r}}_{\mathbf{x}} \in \mathcal{B}_{\mathsf{d}}(\mathbf{r_x}, \ell)$ from the noisy observation $\mathbf{y}$. Fig. 1 provides a graphical representation of the *approximate* ranking recovery problem. We note that due

[1] There exists a one to one mapping between permutation and ranking. In particular, the mapping is $\pi_{\{\cdot\}}$, and it holds that $\pi_{\pi_{\mathbf{x}}} = \mathbf{r_x}$ and $\pi_{\mathbf{r_x}} = \pi_{\mathbf{x}}$.

[2] A random vector $\mathbf{X} \in \mathbb{R}^n$ is said to be exchangeable if $\mathbf{X} \stackrel{d}{=} P\mathbf{X}$ for any permutation matrix P of dimension n.

[3] To be more precise the problem considered in [4], [7] is that of permutation recovery. However, ranking recovery and permutation recovery are equivalent under the *exact* recovery constraint.

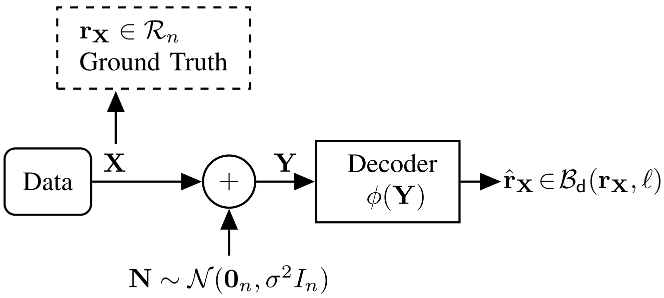

Fig. 1. Graphical representation of the problem framework.

to the assumption **A1**, setting $\ell = 0$ in $\mathcal{B}_{\mathsf{d}}(\mathbf{r}_{\mathbf{x}}, \ell)$ recovers the *exact* version of the problem studied in [4], [7].

We are here interested in analyzing the error probability of the approximate ranking recovery problem, which is given by

$$P_e(\phi, \mathsf{d}, \ell) = \Pr(\mathsf{d}(\mathbf{r}_{\mathbf{X}}, \phi(\mathbf{Y})) > \ell). \tag{5}$$

In particular, our focus will be on characterizing an optimal (i.e., that minimizes (5)) decoder (see Section III), and on understanding how (5) varies with respect to the noise standard deviation $\sigma \in \mathbb{R}_+$ (see Section IV).

III. Optimal Decoder for Approximate Recovery

We here characterize an optimal decoder for the approximate ranking recovery problem, i.e., a decoder that incurs the minimum error probability in (5) in the estimation task. With the definition in (5), an optimal decoder ϕ_{opt} is given by

$$\phi_{\text{opt}} \in \operatorname*{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell), \tag{6}$$

where $P_c(\phi, \mathsf{d}, \ell)$ is the probability of correctness defined as

$$\begin{aligned} P_c(\phi, \mathsf{d}, \ell) &= \Pr(\mathsf{d}(\mathbf{r}_{\mathbf{X}}, \phi(\mathbf{Y})) \leq \ell) \\ &= \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\mathbf{r}_{\mathbf{X}}, \ell)). \end{aligned} \tag{7}$$

As a first result, the following lemma presents a sufficient condition for a decoder ϕ to be optimal.

Lemma 1. *If, for all $\mathbf{y} \in \mathbb{R}^n$, a decoder $\hat{\phi} : \mathbb{R}^n \to \mathcal{R}_n$ satisfies*

$$\hat{\phi}(\mathbf{y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell), \quad \tau = \operatorname*{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y}), \tag{8}$$

then, $\hat{\phi} \in \operatorname{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell)$.

Proof: By using the law of total probability, we can write the probability of correctness in (7) as

$$\begin{aligned} P_c(\phi, \mathsf{d}, \ell) &= \sum_{\tau \in \mathcal{R}_n} \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell), \mathbf{r}_{\mathbf{X}} = \tau) \\ &= \sum_{\tau \in \mathcal{R}_n} \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell) \mid \mathbf{r}_{\mathbf{X}} = \tau)\; p_{\mathbf{r}_{\mathbf{X}}}(\tau) \\ &= \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \Pr(\phi(\mathbf{Y}) = \omega \mid \mathbf{r}_{\mathbf{X}} = \tau)\; p_{\mathbf{r}_{\mathbf{X}}}(\tau) \\ &\overset{(a)}{=} \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \int_{\mathbf{y} \in \mathcal{D}_\omega} f_{\mathbf{Y}|\mathbf{r}_{\mathbf{X}}}(\mathbf{y}|\tau)\; p_{\mathbf{r}_{\mathbf{X}}}(\tau)\; \mathrm{d}\mathbf{y} \\ &\overset{(b)}{=} \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \int_{\mathbf{y} \in \mathcal{D}_\omega} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\tau|\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\, \mathrm{d}\mathbf{y}, \end{aligned} \tag{9}$$

where (a) follows by defining $\mathcal{D}_\omega = \{\mathbf{y} \in \mathbb{R}^n : \phi(\mathbf{y}) = \omega\}$ for all $\omega \in \mathcal{R}_n$, and (b) is due to the Bayes' theorem.

In order to find an optimal decoder ϕ_{opt}, according to (6), we need to maximize $P_c(\phi, \mathsf{d}, \ell)$ with respect to ϕ. Equivalently, with reference to (9), we need to design the decision regions $\mathcal{D}_\omega$'s so as to maximize $P_c(\phi, \mathsf{d}, \ell)$. Towards this end, we note that, for any $\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)$, if we design $\mathcal{D}_\omega$ such that an observation $\mathbf{y} \in \mathcal{D}_\omega$, then the term $p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\tau|\mathbf{y})$ contributes to the integral in (9). For an optimal decoder ϕ_{opt}, we have to guarantee that, for any observation $\mathbf{y} \in \mathbb{R}^n$, the corresponding $\max_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y})$ contributes to (9). It therefore follows that a sufficient condition for a decoder ϕ to be optimal is that, for any observation $\mathbf{y} \in \mathbb{R}^n$ such that $\tau = \operatorname{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y})$, we assign $\mathbf{y}$ to $\mathcal{D}_\omega$, where $\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)$. This concludes the proof of Lemma 1. ■

By leveraging Lemma 1, we are now ready to prove our first main result, which shows that the linear decoder $\phi_{\text{lin}}(\mathbf{y}) = \mathbf{r}_{\mathbf{y}}$ is indeed optimal for the approximate recovery problem.

Theorem 1. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, and suppose that the assumption* **A1** *holds. Then, for any $\ell \geq 0$, we have that*

$$\phi_{\text{lin}} \in \operatorname*{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell). \tag{10}$$

Proof: We consider the Maximum a Posteriori (MAP) decision rule [24], i.e.,[4]

$$\phi_{\text{MAP}}(\mathbf{y}) = \operatorname*{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y}). \tag{11}$$

We note that the assumption **A1** implies that $\omega \in \mathcal{B}_{\mathsf{d}}(\omega, \ell)$ for all $\omega \in \mathcal{R}_n$. Thus, under this assumption, the sufficient conditions in (8) in Lemma 1 are satisfied and hence, it is guaranteed that $\phi_{\text{MAP}} \in \operatorname{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell)$, i.e., ϕ_{MAP} is an optimal decoder. In [4], [7] we showed that, for any exchangeable $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, we have $\phi_{\text{lin}} = \phi_{\text{MAP}}$. This readily implies that $\phi_{\text{lin}} \in \operatorname{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell)$, and concludes the proof of Theorem 1. ■

Remark 1. *The assumption* **A1** *in Theorem 1 for the optimality of the linear decoder is very mild and is known as the identity of indiscernibles. The assumption* **A2** *is also mild. We indeed note that widely adopted distance functions, such as the Hamming distance in Definition 2 and the Kendall's tau rank distance in Definition 3 satisfy these conditions.*

IV. $P_e(\phi_{\text{lin}}, \mathsf{d}, \ell)$ versus σ

Theorem 1 shows the optimality (in terms of error probability) of the linear decoder. In this section, we study the probability of error incurred by such a linear decoder, namely $P_e(\phi_{\text{lin}}, \mathsf{d}, \ell)$, as a function of the noise standard deviation $\sigma \in \mathbb{R}_+$. In particular, different from the *exact* ranking recovery problem (where in the low-noise regime, the probability

[4]We note that $\phi_{\text{MAP}}(\mathbf{y})$ in (11) might not be unique; if this is the case, then we randomly select one of these possible choices.

of error is linear in σ [7]), we show that for the *approximate* version of the problem $P_e(\phi_{\rm lin}, \mathsf{d}, \ell)$ exhibits a sublinear behavior in σ in the low-noise regime (see Theorem 2). This result is also shown in Fig. 2 (which was obtained by using Monte Carlo simulations with 10^6 iterations), and it highlights that relaxing the constraint of *exact* recovery indeed leads to a significantly less noise-dominated problem.

We next introduce and define a few quantities that we will need in the proof of our result. In particular, we borrow the notation and definitions that we have introduced in [7].

Definition 4. *Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector. The i-th order statistics [25] of $\mathbf{X}$ (i.e., the i-th smallest value of $\mathbf{X}$) satisfies*

$$X_{i:n} : X_{1:n} \leq ... \leq X_{i:n} \leq ... \leq X_{n:n}. \qquad (12)$$

Then, we say that the i-th spacing of $\mathbf{X}$ is [26]

$$W_i = X_{i+1:n} - X_{i:n}. \qquad (13)$$

We now state the following lemma, the proof of which can be found in Appendix A.

Lemma 2. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable, $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, and $\tau = (1, 2, \cdots, n)$. Assume that $f_{W_i}(w) < \infty$, $\forall w$, where W_i is defined in (13). Then,*

$$\lim_{\sigma \to 0} \sum_{i=1}^{n-1} \frac{\Pr\left(\mathbf{r_Y} = P^{(i,i+1)}\tau \mid \mathbf{r_X} = \tau\right)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}},$$

where $P^{(i,j)}$ is the permutation matrix of dimension n that permutes the i-th and j-th rankings.

By leveraging Lemma 2, we can now prove the following theorem, which is the second main result of the paper.

Theorem 2. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume that $f_{W_i}(w) < \infty$, $\forall w$ where W_i is defined in (13). Consider a distance function d satisfying the assumptions* **A1** *and* **A2***, and let*

$$\mathsf{d}(\tau, P^{(i,i+1)}\tau) = \beta_i, \ \tau = (1, ..., n), \ \forall i \in [1 : n-1]. \qquad (14)$$

Then, if $\ell \geq \beta^\star = \max_i\{\beta_i\}$, it holds that

$$\lim_{\sigma \to 0} \frac{1}{\sigma} P_e(\phi_{\rm lin}, \mathsf{d}, \ell) = 0. \qquad (15)$$

Proof: We start by observing that

$$\begin{aligned}
&\Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k) \\
&= \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \eta) \\
&\overset{(a)}{=} \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(P_{\tau,\eta}\mathbf{r_X}, P_{\tau,\eta}\mathbf{r_Y}) = k, P_{\tau,\eta}\mathbf{r_X} = \eta) \\
&\overset{(b)}{=} \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \tau) \\
&= n! \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \tau) \\
&= \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau), \qquad (16)
\end{aligned}$$

where the labeled equalities follow from: (a) the fact that $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \mathbf{X} + \mathbf{N}) \overset{d}{=} (P\mathbf{X}, P\mathbf{X} + P\mathbf{N}) = (P\mathbf{X}, P\mathbf{Y})$ for any permutation matrix P due to the exchangeability of $\mathbf{X}$ and $\mathbf{N}$, and letting $P_{\tau,\eta}$ be the permutation matrix that permutes τ into η; and (b) the assumption **A2** and the fact that $P_{\eta,\tau}P_{\tau,\eta} = I_n$ and $P_{\eta,\tau}\eta = \tau$. By using (16) we then obtain

$$\begin{aligned}
&P_e(\phi_{\rm lin}, \mathsf{d}, 0) = \sum_{k>0} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k) \\
&= \sum_{k>0} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&\overset{(a)}{=} \sum_{0<k\leq\beta^\star} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&\quad + \sum_{k>\beta^\star} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&= \sum_{0<k\leq\beta^\star} \Pr(\mathsf{d}(\tau, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) + P_e(\phi, \mathsf{d}, \beta^\star) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n-1} \Pr\left(\mathbf{r_Y} = P^{(i,i+1)}\tau \,\middle|\, \mathbf{r_X} = \tau\right) + P_e(\phi, \mathsf{d}, \beta^\star), \quad (17)
\end{aligned}$$

where (a) follows by letting $\beta^\star = \max_{i\in[1:n-1]}\{\beta_i\}$, and (b) is due to (14).

From [7, Theorem 2], we know that

$$\lim_{\sigma \to 0} \frac{1}{\sigma} P_e(\phi_{\rm lin}, \mathsf{d}, 0) = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}},$$

and from Lemma 2, we have

$$\lim_{\sigma \to 0} \frac{1}{\sigma} \sum_{i=1}^{n-1} \Pr\left(\mathbf{r_Y} = P^{(i,i+1)}\tau \,\middle|\, \mathbf{r_X} = \tau\right) = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}.$$

Thus, the two facts above, together with (17), imply that $\lim_{\sigma\to 0} \frac{1}{\sigma} P_e(\phi_{\rm lin}, \mathsf{d}, \beta^\star) = 0$. We conclude the proof of Theorem 2 by noting that for any $\ell \geq \beta^\star$, we have that $P_e(\phi_{\rm lin}, \mathsf{d}, \ell) \leq P_e(\phi_{\rm lin}, \mathsf{d}, \beta^\star)$, which implies $\lim_{\sigma\to 0} \frac{1}{\sigma} P_e(\phi_{\rm lin}, \mathsf{d}, \ell) = 0$ for all $\ell \geq \beta^\star$. ■

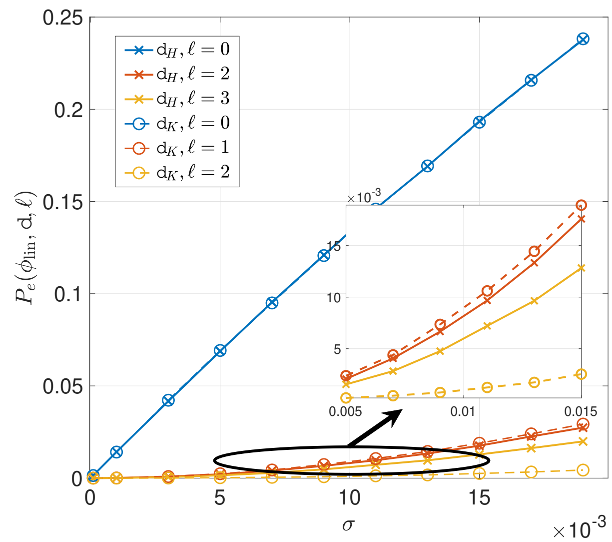

Fig. 2. $P_e(\phi_{\rm lin}, \mathsf{d}, \ell)$ in (5) versus σ with $\mathsf{d} \in \{\mathsf{d}_H, \mathsf{d}_K\}$ and $\ell \in \{0, 1, 2, 3\}$. We set $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_{10}, I_{10})$ and $\mathcal{N} \sim \mathcal{N}(\mathbf{0}_{10}, \sigma^2 I_{10})$.

Remark 2. *The $1/\sigma$ in Theorem 2 is used to prove the sublinear behavior of P_e in the low-noise regime (i.e., the limit in (15) is indeed zero). Theorem 2 implies that in the low-noise regime, errors occur dominantly by interchanging the two entries that are neighbors in terms of ranking. This is because for any $\tau \in \mathcal{R}_n$, the region $\mathcal{H}_\tau = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{r}_\mathbf{x} = \tau\}$ has the $n-1$ regions $\mathcal{H}_\eta$ with $\eta = P^{(i,i+1)}\tau,\ i \in [1:n-1]$, as neighbors.*

We conclude this section with two corollaries on the two practically relevant distances in Definition 2 and Definition 3.

Corollary 1. *For any $\ell \geq 2$, we have that*

$$\lim_{\sigma \to 0} \frac{P_e(\phi_{\text{lin}}, \mathsf{d}_H, \ell)}{\sigma} = 0.$$

Proof: For any $i \in [1:n-1]$ and $\tau \in \mathcal{R}_n$, we have that $\mathsf{d}_H(\tau, P^{(i,i+1)}\tau) = 2 = \beta^\star$. Thus, for any $\ell \geq \beta^\star = 2$, we have that (15) in Theorem 2 holds. This concludes the proof of Corollary 1. ■

Corollary 2. *For any $\ell \geq 1$, we have that*

$$\lim_{\sigma \to 0} \frac{P_e(\phi_{\text{lin}}, \mathsf{d}_K, \ell)}{\sigma} = 0.$$

Proof: The Kendall's tau rank distance satisfies the assumptions **A1** and **A2** and has $\beta^\star = 1$. Hence, from Theorem 2, for any $\ell \geq 1$ we have that (15) in Theorem 2 holds. This concludes the proof of Corollary 2. ■

Appendix A
Proof of Lemma 2

We let $\mathcal{E}_i(\mathbf{Y}) \triangleq \{Y_{i-1} \leq Y_{i+1}\} \cap \{Y_{i+1} \leq Y_i\} \cap \{Y_i \leq Y_{i+2}\}$, and $\mathcal{I}_i \triangleq [1:n-1] \setminus \{i-1, i, i+1\}$. We have,

$$\begin{aligned}
&\Pr\left(\mathbf{r}_\mathbf{Y} = P^{(i,i+1)}\tau \,\middle|\, \mathbf{r}_\mathbf{X} = \tau\right) \\
&\stackrel{(a)}{=} \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{Y_t \leq Y_{t+1}\} \cap \mathcal{E}_i(\mathbf{Y}) \,\middle|\, \tau\right) \\
&\stackrel{(b)}{=} \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{V_t \leq W_t\} \cap \mathcal{E}_i(\mathbf{X}+\mathbf{N}) \,\middle|\, \tau\right) \\
&\stackrel{(c)}{=} \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{V_t \leq W_t\} \cap \tilde{\mathcal{E}}_i(\mathbf{X}+\mathbf{N}) \,\middle|\, \tau\right) \\
&\stackrel{(d)}{=} \Pr\left(V_i \leq -W_i\right) \\
&\times \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{V_t \leq W_t\} \cap \mathcal{E}_i^\star(\mathbf{V}, \mathbf{W}) \,\middle|\, V_i \leq -W_i\right),
\end{aligned} \quad (18)$$

where the labeled equalities follow from: (a) the fact that $\tau = (1, 2, \cdots, n)$, and letting $\Pr(\cdot|\tau) = \Pr(\cdot|\mathbf{r}_\mathbf{X} = \tau)$ for brevity; (b) Definition 4 for which $W_t = X_{t+1} - X_t$ and letting $V_t = N_t - N_{t+1}$; note that, with reference to Definition 4 we have that $X_{i:n} \stackrel{d}{=} X_i$ given the condition $\mathbf{r}_\mathbf{X} = \tau$; (c) noting that, since $\mathbf{N}$ is exchangeable, the event $\mathcal{E}_i(\mathbf{X}+\mathbf{N})$ is equal in distribution to the event $\tilde{\mathcal{E}}_i(\mathbf{X}+\mathbf{N})$ given as follows,

$$\begin{aligned}
\tilde{\mathcal{E}}_i(\mathbf{X}+\mathbf{N}) &\stackrel{(c1)}{=} \{X_{i-1} + N_{i-1} \leq X_{i+1} + N_i\} \\
&\quad \cap \{X_{i+1} + N_i \leq X_i + N_{i+1}\} \\
&\quad \cap \{X_i + N_{i+1} \leq X_{i+2} + N_{i+2}\} \\
&\stackrel{(c2)}{=} \{V_{i-1} \leq W_{i-1} + W_i\} \cap \{V_i \leq -W_i\} \\
&\quad \cap \{V_{i+1} \leq W_i + W_{i+1}\},
\end{aligned}$$

where (c1) follows by permuting N_i and N_{i+1}, and (c2) follows since $\mathbf{r}_\mathbf{X} = \tau$ and by using Definition 4 for $W_t = X_{t+1} - X_t$ and $V_t = N_t - N_{t+1}$; and (d) introducing $\mathcal{E}_i^\star(\mathbf{V}, \mathbf{W}) \triangleq \{V_{i-1} \leq W_{i-1} + W_i\} \cap \{V_{i+1} \leq W_i + W_{i+1}\}$, and using the definition of conditional probability.

We now analyze the two probability terms in (18). The first probability term in (18) is

$$\begin{aligned}
\Pr\left(V_i \leq -W_i\right) &= \Pr\left(Z \leq -\frac{W_i}{\sigma\sqrt{2}}\right) \\
&= \int_0^\infty Q\left(\frac{w}{\sigma\sqrt{2}}\right) f_{W_i}(w)\, \mathrm{d}w \\
&= \int_0^\infty Q\left(u\right) f_{W_i}(\sqrt{2}\sigma u)\sqrt{2}\sigma\, \mathrm{d}u,
\end{aligned} \quad (19)$$

where $Q(\cdot)$ is the standard Gaussian Q function, and the last equality follows by a change of variable. By dividing (19) by σ and taking $\sigma \to 0$, we obtain

$$\begin{aligned}
&\lim_{\sigma \to 0} \frac{\Pr\left(V_i \leq -W_i\right)}{\sigma} \stackrel{(a)}{=} \int_0^\infty Q\left(u\right) \lim_{\sigma \to 0} f_{W_i}(\sqrt{2}\sigma u)\sqrt{2}\, \mathrm{d}u \\
&= \int_0^\infty Q\left(u\right) f_{W_i}(0^+)\sqrt{2}\, \mathrm{d}u = \frac{f_{W_i}(0^+)}{\sqrt{\pi}},
\end{aligned} \quad (20)$$

where (a) follows from the dominated convergence theorem, which is verifiable since $f_{W_i}(w) \leq \sup f_{W_i}(w) < \infty$, and $\int_0^\infty Q(u)\, \mathrm{d}u$ is integrable.

For the second probability term in (18) we have,

$$\begin{aligned}
&\lim_{\sigma \to 0} \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{V_t \leq W_t\} \cap \mathcal{E}_i^\star(\mathbf{V}, \mathbf{W}) \,\middle|\, V_i \leq -W_i\right) \\
&\stackrel{(a)}{=} \lim_{\sigma \to 0} \Pr\left(\bigcap_{t \in \mathcal{I}_i} \{\sigma\tilde{V}_t \leq W_t\} \cap \mathcal{E}_i^\star(\sigma\tilde{\mathbf{V}}, \mathbf{W}) \,\middle|\, V_i \leq -W_i\right) \\
&\stackrel{(b)}{=} 1,
\end{aligned} \quad (21)$$

where (a) follows by letting $\mathbf{V} = \sigma\tilde{\mathbf{V}}$ with $\tilde{V}_t = \frac{1}{\sigma}(N_t - N_{t+1})$, and (b) is due to the fact that $\mathbf{W} \geq \mathbf{0}_{n-1}$.

By using (18), (20) and (21), we obtain

$$\lim_{\sigma \to 0} \frac{\Pr\left(\mathbf{r}_\mathbf{Y} = P^{(i,i+1)}\tau \mid \mathbf{r}_\mathbf{X} = \tau\right)}{\sigma} = \frac{f_{W_i}(0^+)}{\sqrt{\pi}},$$

and hence,

$$\lim_{\sigma \to 0} \sum_{i=1}^{n-1} \frac{\Pr\left(\mathbf{r}_\mathbf{Y} = P^{(i,i+1)}\tau \mid \mathbf{r}_\mathbf{X} = \tau\right)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}.$$

This concludes the proof of Lemma 2.

REFERENCES

[1] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the web," in *Proceedings of the 10th International Conference on World Wide Web*, 2001, pp. 613–622.

[2] S. Chanas and P. Kobylański, "A new heuristic algorithm solving the linear ordering problem," *Computational Optimization and Applications*, vol. 6, no. 2, pp. 191–205, 1996.

[3] J. P. Baskin and S. Krishnamurthi, "Preference aggregation in group recommender systems for committee decision-making," in *Proceedings of the third ACM Conference on Recommender Systems*, 2009, pp. 337–340.

[4] M. Jeong, A. Dytso, M. Cardone, and H. V. Poor, "Recovering data permutations from noisy observations: The linear regime," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 854–869, 2020.

[5] S. Özyurt and O. Kucur, "Power permutation modulation in multiple-input multiple-output systems," *Transactions on Emerging Telecommunications Technologies*, p. e4408, 2021.

[6] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy." *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.

[7] M. Jeong, A. Dytso, and M. Cardone, "Retrieving data permutations from noisy observations: High and low noise asymptotics," in *Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT)*, July 2021.

[8] S. R. Searle *et al.*, "Prediction, mixed models, and variance components," 1973.

[9] S. Portnoy, "Maximizing the probability of correctly ordering random variables using linear predictors," *Journal of Multivariate Analysis*, vol. 12, no. 2, pp. 256–269, 1982.

[10] K. Nomakuchi and T. Sakata, "Characterizations of the forms of covariance matrix of an elliptically contoured distribution," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 205–210, 1988.

[11] ——, "Characterization of conditional covariance and unified theory in the problem of ordering random variables," *Annals of the Institute of Statistical Mathematics*, vol. 40, no. 1, pp. 93–99, 1988.

[12] O. Collier and A. S. Dalalyan, "Minimax rates in permutation estimation for feature matching," *The Journal of Machine Learning Research*, vol. 17, no. 6, pp. 1 –31, January 2016.

[13] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3286–3300, May 2018.

[14] ——, "Denoising linear models with permuted data," in *Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 446–450.

[15] P. Rigollet and J. Weed, "Uncoupled isotonic regression via minimum Wasserstein deconvolution," *Information and Inference: A Journal of the IMA*, vol. 8, no. 4, pp. 691–717, December 2019.

[16] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3237–3253, May 2018.

[17] S. Haghighatshoar and G. Caire, "Signal recovery from unlabeled samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1242–1257, March 2018.

[18] H. Zhang, M. Slawski, and P. Li, "Permutation recovery from multiple measurement vectors in unlabeled sensing," in *Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1857–1861.

[19] I. Dokmanić, "Permutations unlabeled beyond sampling unknown," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 823–827, April 2019.

[20] M. Tsakiris and L. Peng, "Homomorphic sensing," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, June 2019, pp. 6335–6344.

[21] M. C. Tsakiris, "Eigenspace conditions for homomorphic sensing," *arXiv:1812.07966*, April 2019.

[22] D. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*, December 2017, pp. 1530–1539.

[23] A. Dytso, M. Cardone, M. S. Veedu, and H. V. Poor, "On estimation under noisy order statistics," in *Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 36–40.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing, vol. 2: Detection Theory*. Prentice Hall PTR, 1998.

[25] H. A. David and H. N. Nagaraja, "Order statistics," *Encyclopedia of statistical sciences*, 2004.

[26] R. Pyke, "Spacings," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 27, no. 3, pp. 395–449, 1965. [Online]. Available: http://www.jstor.org/stable/2345793