Recovering Structure of Noisy Data through Hypothesis Testing

Minoh Jeong
University of Minnesota
Minneapolis, MN 55404, USA
Email: jeong316@umn.edu

Alex Dytso
Princeton University
Princeton, NJ 08544, USA
Email: adytso@princeton.edu

Martina Cardone
University of Minnesota
Minneapolis, MN 55404, USA
Email: mcardone@umn.edu

H. Vincent Poor
Princeton University
Princeton, NJ 08544, USA
Email: poor@princeton.edu

***Abstract*—This paper considers a noisy data structure recovery problem. Specifically, the goal is to investigate the following question: Given a noisy observation of the data, according to which permutation was the original data sorted? The main focus is on scenarios where data is generated according to an isotropic Gaussian distribution, and the perturbation consists of adding Gaussian noise with diagonal scalar covariance matrix. This problem is posed within a hypothesis testing framework. First, the optimal decision criterion is characterized and shown to be identical to the hypothesis of the observation. Then, by leveraging the structure of the optimal decision criterion, the probability of error is characterized. Finally, the logarithmic behavior (i.e., the exponent) of the probability of error is derived in the regime where the dimension of the data goes to infinity.**

I. Introduction

The problem of recovering data structure, given a perturbed observation of it, is becoming a prevailing task of modern communication and computing systems. For instance, in a recommender system, users may desire to *privatize* their data before it is collected from an external party. A suitable solution to privatize data, and hence maintain its confidentiality, consists of perturbing it with some noise. Upon receiving the perturbed/noisy data the recommender system might need to recover the data structure (e.g., ranking of users' interests) in order to provide the next recommendation.

In this work, we are interested in investigating the following question on noisy data structure recovery: Given a noisy observation of the data, according to which permutation was the original data sorted? In particular, we consider a scenario where data is generated according to a Gaussian distribution, and the perturbation consists of adding Gaussian noise. The main focus is on Gaussian noise perturbations for which the covariance matrix is diagonal with all elements equal to σ^2.

We start our analysis by formulating the problem within a hypothesis testing framework, which consists of $n!$ hypotheses, where n is the dimension of the data vector. We then characterize the optimal decision criterion for the hypothesis testing problem, by deriving the optimal decision regions. In particular, we show that the optimal decision is identical to the hypothesis of the observation, and is independent of σ^2.

The work of M. Jeong and M. Cardone was supported in part by the U.S. National Science Foundation under Grant CCF-1849757. The work of A. Dytso and H. V. Poor was supported in part by the U.S. National Science Foundation under Grant CCF-1908308

With the structure of the optimal decision regions, we finally proceed to characterize the probability of error incurred by the decision criterion. In particular, we express the probability of error in terms of the volume of a region which consists of the intersection of a cone with a linear transformation of the unit radius n-dimensional ball. We also use this characterization to derive the logarithmic behavior (i.e., the exponent) of the probability of error as n goes to infinity.

Related Work. The problem of permutation recovery has recently gained significant importance, and it is widely studied in several fields, such as computer science and (bio)statistics.

In the machine learning literature, a problem of permutation estimation given noisy observations is studied in [1]. The model in [1] consists of recovering the permutation needed to match two sets of features, given noisy observations of them. In particular, the authors provide a separation rate parameter for estimating the permutation, and a minimax separation rate for recovering the permutation. In [2] the authors analyze a framework for estimating not only the structure/permutation, but also the values of an original sorted vector perturbed by noise, by performing joint estimation and sorting.

A permutation recovery problem also appears in linear regression [3], [4]. The authors consider an additive noise linear regression model in which the output is permuted by an unknown permutation matrix. In [3], conditions for (approximate) permutation recovery are discussed. Under the same model, a characterization of the minimax prediction error and estimators are discussed in [4]. In [5], multivariate linear regression under sparse permutation is considered. The goal is to recover the permutation, which acts only on part of the data. Other interesting works on this topic are [6]–[8]. The uncoupled isotonic regression problem, where the goal consists of estimating a non-decreasing regression function given unordered sets of data, is studied in [9].

A study on permutation recovery in biostatistics is found in [10]. The authors discuss exact and partial permutation recoveries under Kendall tau distance for the microbiome growth dynamics. Further, an interesting binary hypothesis testing detection problem with unknown permutation, namely unlabeled detection, is discussed in [11].

Paper Organization. Section II introduces the notation and formulates the hypothesis testing problem. Section III discusses the optimal decision regions for our hypothesis testing

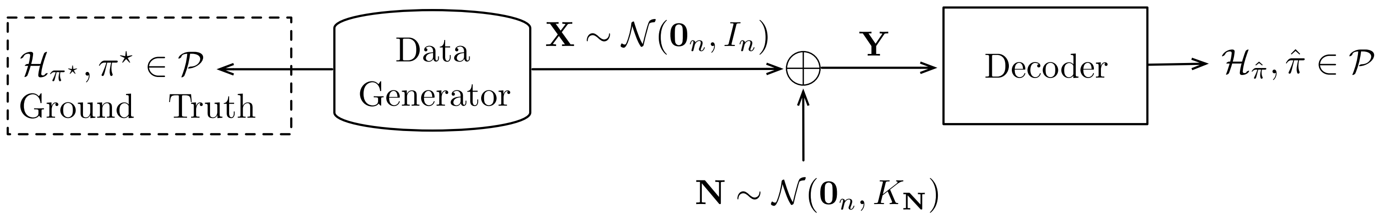

Fig. 1. Graphical representation of the proposed framework.

problem. Section IV presents a new characterization for the probability of error and illustrates its logarithmic behavior as n goes to infinity. Section V concludes the paper.

II. Notation and Problem Formulation

Notation. Boldface upper case letters $\mathbf{X}$ denote vector 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; $0_{n\times n}$ is the zero matrix of dimension $n \times n$; A^{-1} is the inverse of the square matrix A; $\|A\|$ is the spectral norm of the matrix A and $\det(A)$ is the determinant; $\|\mathbf{x}\|$ is the ℓ_2 norm of $\mathbf{x}$, and $\mathbf{x}^T$ is the transpose of $\mathbf{x}$. Calligraphic letters indicate sets; $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$; for two sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \setminus \mathcal{B}$ is the set of elements that belong to $\mathcal{A}$ but not to $\mathcal{B}$, $\mathcal{A} \cap \mathcal{B}$ is the set of elements that belong both to $\mathcal{A}$ and $\mathcal{B}$, and $\mathcal{A} \cup \mathcal{B}$ is the set of elements which are in either set; $\varnothing$ is the empty set.

We consider the framework in Fig. 1, where an n-dimensional random vector $\mathbf{X}$ is generated according to an isotropic Gaussian probability density function (PDF), namely $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. The vector $\mathbf{X}$ is passed through an additive noisy channel with Gaussian transition probability, the output of which is denoted as $\mathbf{Y}$. Thus, we have $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, with $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ where $K_{\mathbf{N}}$ denotes the covariance matrix of the additive noise $\mathbf{N}$, and where $\mathbf{X}$ and $\mathbf{N}$ are independent.

In this work, we are interested in answering the following question: Given the observation of $\mathbf{Y}$, according to which permutation - among the $n!$ possible ones - was the vector $\mathbf{X}$ sorted? Towards this end, we define $\mathcal{P}$ as the collection of all permutations of the elements of $[1 : n]$; clearly $|\mathcal{P}| = n!$. We formulate a hypothesis testing problem with $n!$ hypotheses $\mathcal{H}_\pi, \pi \in \mathcal{P}$, where $\mathcal{H}_\pi$ is the hypothesis that $\mathbf{X}$ is an n-dimensional vector sorted according to the permutation $\pi \in \mathcal{P}$. Formally, each hypothesis corresponds to the following set,

$$\mathcal{H}_\pi = \{\mathbf{x} : x_{\pi_1} \leq x_{\pi_2} \leq \cdots \leq x_{\pi_n}\}, \tag{1}$$

where $x_{\pi_i}, i \in [1 : n]$ is the π_i-th element of $\mathbf{x}$, and $\pi_i, i \in [1 : n]$ is the i-th element of π.

We seek to characterize the *optimal* decision criterion among the $n!$ hypotheses, as well as assessing its performance in terms of error probability. In other words, with reference to Fig. 1, we are interested in characterizing the *decision criterion* so that its output $\mathcal{H}_{\hat{\pi}}, \hat{\pi} \in \mathcal{P}$ is such that

$$\mathcal{H}_{\hat{\pi}} : \ \hat{\pi} = \underset{\pi \in \mathcal{P}}{\operatorname{argmin}} \left\{\Pr\left(\mathcal{H}_\pi \neq \mathcal{H}_{\pi^\star}\right)\right\}, \tag{2}$$

where $\pi^\star$ denotes the permutation according to which the random vector $\mathbf{X}$ is sorted.

Example. Let $n = 3$, then we have $|\mathcal{P}| = 6$ and hypotheses $\mathcal{H}_\pi, \pi \in \mathcal{P}$ defined as

$$\begin{aligned} &\mathcal{H}_{\{1,2,3\}} : X_1 \leq X_2 \leq X_3, \quad &\mathcal{H}_{\{1,3,2\}} : X_1 \leq X_3 \leq X_2,\\ &\mathcal{H}_{\{2,1,3\}} : X_2 \leq X_1 \leq X_3, \quad &\mathcal{H}_{\{2,3,1\}} : X_2 \leq X_3 \leq X_1,\\ &\mathcal{H}_{\{3,1,2\}} : X_3 \leq X_1 \leq X_2, \quad &\mathcal{H}_{\{3,2,1\}} : X_3 \leq X_2 \leq X_1, \end{aligned}$$

where $X_i, i \in [1 : 3]$ is the i-th element of $\mathbf{X}$. Each hypothesis is hence associated to a region – referred to as *hypothesis region* – in the 3-dimensional space, as also graphically represented in Fig. 2.

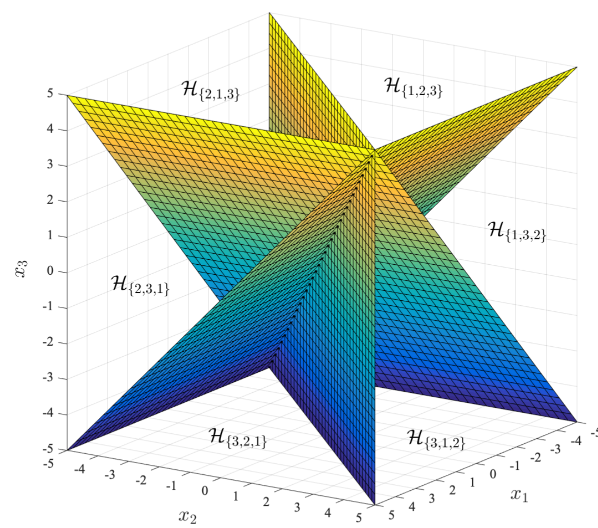

Fig. 2. Case $n = 3$. Graphical representation of the hypothesis regions associated to each of the 6 hypotheses.

III. Optimal Decision Regions

In this section, we seek to solve the optimization problem in (2), hence characterizing the optimal decision criterion. Towards this end, we make use of the result in [12, Appendix 3C], which shows that, for an observed $\mathbf{y}$, the optimal decision criterion in (2) is given by the maximum a posterior probability (MAP) decoder, namely

$$\mathcal{H}_{\hat{\pi}} : \ \hat{\pi} = \underset{\pi \in \mathcal{P}}{\operatorname{argmax}} \left\{f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\pi)\right\}, \tag{3a}$$

$$f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\pi) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\pi) \Pr(\mathcal{H}_\pi), \ \pi \in \mathcal{P}. \tag{3b}$$

By defining the likelihood functions $L(\mathbf{y}, \mathcal{H}_\pi) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\pi), \forall \pi \in \mathcal{P}$, we have that (3) can be equivalently formulated as

$$\mathcal{H}_{\hat{\pi}} : \frac{L(\mathbf{y}, \mathcal{H}_{\hat{\pi}})}{L(\mathbf{y}, \mathcal{H}_\pi)} > 1, \ \forall \pi \neq \hat{\pi}, \tag{4}$$

where we have used the fact that $\Pr(H_\pi) = \Pr(H_\tau), \forall (\pi, \tau) \in \mathcal{P} \times \mathcal{P}$, which follows since $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. It is worth noting that, since $\mathbf{X}$ and $\mathbf{N}$ are independent, then the likelihood function $L(\mathbf{y}, \mathcal{H}_\pi), \pi \in \mathcal{P}$ can be expressed by using the convolution between two PDFs as

$$L(\mathbf{y}, \mathcal{H}_\pi) = \mathbb{E}\left[f_{\mathbf{N}}(\mathbf{y} - \mathbf{X}) | \mathcal{H}_\pi\right], \tag{5}$$

where $f_{\mathbf{N}}(\cdot)$ is the PDF of $\mathbf{N}$ and $*$ indicates the convolution operation.

With the formulation in (4), we can now define the *optimal* decision regions $\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{P}$ of our hypothesis testing problem[1]. In particular, the decision criterion will leverage these regions to output $\mathcal{H}_{\hat{\pi}}, \hat{\pi} \in \mathcal{P}$, namely if the observation vector $\mathbf{y} \in \mathcal{R}_{\pi, K_{\mathbf{N}}}$, then the decoder declares that the input vector $\mathbf{x} \in \mathcal{H}_\pi$. We have that the optimal decision region $\mathcal{R}_{\pi, K_{\mathbf{N}}}$ corresponding to the hypothesis $\mathcal{H}_\pi, \pi \in \mathcal{P}$ is defined as

$$\begin{aligned}\mathcal{R}_{\pi, K_{\mathbf{N}}} &= \left\{ \mathbf{y} \in \mathbb{R}^n : f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\pi) > \max_{\substack{\tau \in \mathcal{P} \\ \tau \neq \pi}} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) \right\} \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{L(\mathbf{y}, \mathcal{H}_\pi)}{L(\mathbf{y}, \mathcal{H}_\tau)} > 1, \ \forall \tau \in \mathcal{P}, \tau \neq \pi \right\}. \end{aligned} \tag{6}$$

We highlight that each decision region described in (6) is a set of observation vectors $\mathbf{y} \in \mathbb{R}^n$ that satisfy the optimal decision criterion in (4).

Remark 1. *In order to guarantee that the collection $\{\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{P}\}$ is a partition of the n-dimensional space, we assume that if the observation vector $\mathbf{y} \in \mathbb{R}^n$ belongs to the boundary between two or more decision regions, i.e., $\mathbf{y} \in \{\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{S}, \mathcal{S} \subseteq \mathcal{P}, |\mathcal{S}| > 1\}$, then we arbitrarily select one of the hypotheses $\mathcal{H}_\pi, \pi \in \mathcal{S}$.*

We next present one of our main results, which fully characterizes the optimal decision regions in (6) for the case when $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ with $K_{\mathbf{N}} = \sigma^2 I_n$. Under this assumption, the regions $\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{P}$ depend on $K_{\mathbf{N}}$ only through the parameter σ. Hence, in what follows we let $\mathcal{R}_{\pi, K_{\mathbf{N}}} = \mathcal{R}_{\pi, \sigma}$.

Theorem 1. *Assume that $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ with $K_{\mathbf{N}} = \sigma^2 I_n$. Then, the optimal decision regions in* (6) *are given by*

$$\mathcal{R}_{\pi, \sigma} = \mathcal{H}_\pi, \ \forall \pi \in \mathcal{P}. \tag{7}$$

Before delving into the proof of Theorem 1, we state two remarks that highlight some properties of the result in (7).

Remark 2. *We note that the result in Theorem 1 can be extended beyond Gaussian $\mathbf{X}$ and holds for any $\mathbf{X}$ that has a spherically symmetric distribution, i.e., $\mathbf{X}$ and $U\mathbf{X}$ have the same distribution for every unitary matrix U.*

[1]The notation $\mathcal{R}_{\pi, K_{\mathbf{N}}}$ indicates that, in general, the decision regions might be functions of the noise covariance matrix $K_{\mathbf{N}}$.

Remark 3. *The result in* (7) *shows that when $K_{\mathbf{N}} = \sigma^2 I_n$, then the optimal decision is identical to the hypothesis of observation. This result implies that: (i) there is no need of performing any further operation, such as likelihood ratio comparisons, to obtain the optimal decision criterion; and (ii) the optimal decision is independent of the value of σ. These results are a consequence of the symmetry of $\mathbf{X}$ and $\mathbf{N}$.*

In the remaining of this section, we focus on proving the result in Theorem 1. In particular, we use the notation $\mathbf{h}_\pi$ with $\pi \in \mathcal{P}$ to denote the fact that the components of $\mathbf{h}_\pi$ are sorted according to the permutation π, i.e., $\mathbf{h}_\pi \in \mathcal{H}_\pi$. We now note that for any vector $\mathbf{h}_\pi, \pi \in \mathcal{P}$, we have that

$$f_{\mathbf{Y}}(\mathbf{h}_\pi | \mathcal{H}_\pi) = \frac{1}{\Pr(\mathcal{H}_\pi)} \int_{\mathbf{x}_\pi \in \mathcal{H}_\pi} f_{\mathbf{N}}(\mathbf{h}_\pi - \mathbf{x}_\pi) f_{\mathbf{X}}(\mathbf{x}_\pi) \ \mathrm{d}\mathbf{x}_\pi, \tag{8}$$

and similarly for $\tau \in \mathcal{P}$ with $\tau \neq \pi$ we have that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{h}_\pi | \mathcal{H}_\tau) &= \frac{1}{\Pr(\mathcal{H}_\tau)} \int_{\mathbf{x}_\tau \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{h}_\pi - \mathbf{x}_\tau) f_{\mathbf{X}}(\mathbf{x}_\tau) \ \mathrm{d}\mathbf{x}_\tau \\ &\overset{(a)}{=} \frac{1}{\Pr(\mathcal{H}_\pi)} \int_{\mathbf{x}_\pi \in \mathcal{H}_\pi} f_{\mathbf{N}}(\mathbf{h}_\pi - P_{\pi,\tau}\mathbf{x}_\pi) f_{\mathbf{X}}(P_{\pi,\tau}\mathbf{x}_\pi) \ \mathrm{d}\mathbf{x}_\pi, \end{aligned} \tag{9}$$

where the equality in (a) follows from the fact that, by assumption, we have $\Pr(\mathcal{H}_\pi) = \Pr(\mathcal{H}_\tau)$, and by using a change of variable where $P_{\pi,\tau}$ is the permutation matrix that permutes the vector $\mathbf{x}_\pi \in \mathcal{H}_\pi$ into $P_{\pi,\tau}\mathbf{x}_\pi = \mathbf{x}_\tau \in \mathcal{H}_\tau$.

Thus, from (8) and (9), we obtain

$$\begin{aligned} & f_{\mathbf{Y}}(\mathbf{h}_\pi | \mathcal{H}_\pi) - f_{\mathbf{Y}}(\mathbf{h}_\pi | \mathcal{H}_\tau) \\ &\overset{(a)}{\geq} \int_{\mathbf{x}_\pi \in \mathcal{H}_\pi} \left\{ f_{\mathbf{N}}(\mathbf{h}_\pi - \mathbf{x}_\pi) - f_{\mathbf{N}}(\mathbf{h}_\pi - P_{\pi,\tau}\mathbf{x}_\pi) \right\} f_{\mathbf{X}}(\mathbf{x}_\pi) \ \mathrm{d}\mathbf{x}_\pi \\ &\overset{(b)}{\geq} 0, \end{aligned} \tag{10}$$

where the inequality in (a) follows since $\Pr(\mathcal{H}_\pi) \leq 1$, and the inequality in (b) follows because of the following lemma.

Lemma 1. *Assume $\pi \in \mathcal{P}$ and let $\mathbf{g}_\pi \in \mathcal{H}_\pi$ and $\mathbf{h}_\pi \in \mathcal{H}_\pi$. Then, for all $\tau \in \mathcal{P}$ with $\tau \neq \pi$, we have*

$$\|\mathbf{g}_\pi - \mathbf{h}_\pi\|^2 \leq \|\mathbf{g}_\pi - P_{\pi,\tau}\mathbf{h}_\pi\|^2,$$

where $P_{\pi,\tau}$ is the permutation matrix that permutes the vector $\mathbf{h}_\pi \in \mathcal{H}_\pi$ into $P_{\pi,\tau}\mathbf{h}_\pi = \mathbf{h}_\tau \in \mathcal{H}_\tau$.

Proof: We start by noting that for any two vectors $\mathbf{g}_\pi \in \mathcal{H}_\pi$ and $\mathbf{h}_\tau \in \mathcal{H}_\tau$, where $\pi \in \mathcal{P}$ and $\tau \in \mathcal{P}$, the maximum inner product is obtained when $\pi = \tau$, i.e., when the two vectors are sorted according to the same permutation. Formally, we have

$$\underset{\tau \in \mathcal{P}}{\operatorname{argmax}} \ \{\mathbf{g}_\pi^T \mathbf{h}_\tau\} = \pi. \tag{11}$$

Moreover, for $\pi \in \mathcal{P}$, we have

$$\begin{aligned} \|\mathbf{g}_\pi - \mathbf{h}_\pi\|^2 &= \|\mathbf{g}_\pi\|^2 - 2\mathbf{g}_\pi^T \mathbf{h}_\pi + \|\mathbf{h}_\pi\|^2 \\ &\overset{(a)}{\leq} \|\mathbf{g}_\pi\|^2 - 2\mathbf{g}_\pi^T P_{\pi,\tau}\mathbf{h}_\pi + \|P_{\pi,\tau}\mathbf{h}_\pi\|^2 \\ &= \|\mathbf{g}_\pi - P_{\pi,\tau}\mathbf{h}_\pi\|^2, \end{aligned} \tag{12}$$

where (a) follows from (11) and since $\|\mathbf{h}_\pi\|^2 = \|P_{\pi,\tau}\mathbf{h}_\pi\|^2$. This concludes the proof of Lemma 1. ■

From the result in (10), it follows that, if the observation $\mathbf{y} \in \mathcal{H}_\pi$ with $\pi \in \mathcal{P}$, then $f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\pi) \geq f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau), \forall \tau \in \mathcal{P}, \tau \neq \pi$. This together with the definition in (6) implies that

$$\mathcal{H}_\pi \subseteq \mathcal{R}_{\pi,\sigma}, \ \forall \pi \in \mathcal{P}. \tag{13}$$

The equality follows since both $\mathcal{H}_\pi, \pi \in \mathcal{P}$, and $\mathcal{R}_{\pi,\sigma}, \pi \in \mathcal{P}$, partition the entire n-dimensional space. Formally, assume that, for a permutation $\rho \in \mathcal{P}$, there exists $\mathbf{t} \in \mathcal{R}_{\rho,\sigma} \setminus \mathcal{H}_\rho$. Then, since the $\mathcal{H}_\pi$'s satisfy the condition in (13) and partition the entire n-dimensional space, we must have that there exists $\tau \in \mathcal{P}, \tau \neq \rho$, such that $\mathbf{t} \in \mathcal{R}_{\rho,\sigma} \cap \mathcal{H}_\tau$. However, by the condition in (13) and since also the $\mathcal{R}_{\pi,\sigma}$'s partition the entire n-dimensional space, we have that $\mathcal{R}_{\rho,\sigma} \cap \mathcal{H}_\tau = \varnothing, \forall \rho \neq \tau$. It follows that such a $\mathbf{t}$ cannot exist, which contradicts the assumption. This shows that $\mathcal{H}_\pi = \mathcal{R}_{\pi,\sigma}, \forall \pi \in \mathcal{P}$ and concludes the proof of Theorem 1.

IV. Probability of Error

In this section, we assume $K_{\mathbf{N}} = \sigma^2 I_n$ and we characterize the probability of error incurred by the optimal decision criterion in Theorem 1. In particular, in Section IV-A we express the probability of error in terms of the volume of a region which consists of the intersection of a cone with a linear transformation of the unit radius $2n$-dimensional ball. Then, in Section IV-B we use this characterization to derive the logarithmic behavior (i.e., the exponent) of the probability of error as n goes to infinity.

Next, the n-dimensional ball centered at $\mathbf{c}$ with radius r is denoted as $\mathcal{B}^n(\mathbf{c}, r)$, and the volume (i.e., the n-dimensional Lebesgue measure) of a set $\mathcal{S} \subset \mathbb{R}^n$ is denoted as $\mathrm{Vol}(\mathcal{S})$.

A. Characterization of the Probability of Error

The following theorem characterizes the probability of error incurred by the optimal decision criterion in Theorem 1.

Theorem 2. *Assume that $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ with $K_{\mathbf{N}} = \sigma^2 I_n$. Then, the probability of error is given by*

$$P_e = 1 - n! \frac{\mathrm{Vol}(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}{\sigma^n \mathrm{Vol}(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}, \tag{14}$$

where $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & \sigma I_n \end{bmatrix} \in \mathbb{R}^{2n}$, $\mathcal{C}_{\mathcal{H}_\pi} = \mathcal{H}_\pi \times \mathcal{H}_\pi$, and $\pi \in \mathcal{P}$ can be chosen arbitrarily.

Proof: Instead of working with the probability of error, it is more convenient to work with the probability of correctness of our hypothesis testing problem. Using the structure of the optimal decision regions found in Theorem 1, the probability of correctness can be written as

$$\begin{aligned} P_c &= \sum_{\pi \in \mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{Y})^T \in \mathcal{H}_\pi \times \mathcal{R}_{\pi,\sigma} \right) \\ &= \sum_{\pi \in \mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{Y})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &= n! \Pr\left((\mathbf{X}, \mathbf{Y})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right), \end{aligned} \tag{15}$$

where the last equality follows from the symmetry of $\mathbf{X}$ and $\mathbf{Y}$, and by choosing $\pi \in \mathcal{P}$ arbitrarily.

Furthermore, let $\boldsymbol{Z} \in \mathbb{R}^n$ denote the standard normal random vector, i.e., $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. Then,

$$\begin{aligned} &\Pr\left((\mathbf{X}, \mathbf{Y})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &= \Pr\left((\mathbf{X}, \mathbf{X} + \sigma\boldsymbol{Z})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &= \Pr\left(A(\mathbf{X}, \boldsymbol{Z})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &= \Pr\left((\mathbf{X}, \boldsymbol{Z})^T \in A^{-1}\mathcal{C}_{\mathcal{H}_\pi} \right), \end{aligned} \tag{16}$$

where $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & \sigma I_n \end{bmatrix}$ and $\mathcal{C}_{\mathcal{H}_\pi} = \mathcal{H}_\pi \times \mathcal{H}_\pi$. Note that A is invertible for $\sigma > 0$.

We observe that the shape of the region $\mathcal{H}_\pi$ is an n-dimensional cone (see Fig. 2 for a graphical representation when $n = 3$). Thus, $\mathcal{C}_{\mathcal{H}_\pi}$ is a $2n$-dimensional cone and so is $A^{-1}\mathcal{C}_{\mathcal{H}_\pi}$. It therefore follows that we have to determine the probability of $(\mathbf{X}, \boldsymbol{Z})^T$ to fall within a cone. Using the symmetry of the Gaussian distribution, the probability of a pair $(\mathbf{X}, \boldsymbol{Z})^T$ to fall within a cone is simply determined by the angular measure of the cone. Now, the angular measure of the cone $A^{-1}\mathcal{C}_{\mathcal{H}_\pi}$ is given by

$$\begin{aligned} &\Pr\left((\mathbf{X}, \boldsymbol{Z})^T \in A^{-1}\mathcal{C}_{\mathcal{H}_\pi} \right) \\ &= \frac{\mathrm{Vol}\left(A^{-1}\mathcal{C}_{\mathcal{H}_\pi} \cap \mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)}{\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)} \\ &= \frac{|\det(A^{-1})| \mathrm{Vol}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)}{\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)}, \end{aligned} \tag{17}$$

where in the last equality we have used the fact that $\mathrm{Vol}(A\mathcal{S}) = |\det(A)| \mathrm{Vol}(\mathcal{S})$ for any invertible matrix A and any set $\mathcal{S}$. By combining (15), (16) and (17) we arrive at

$$P_c = n! \frac{|\det(A^{-1})| \mathrm{Vol}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)}{\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1) \right)}. \tag{18}$$

The proof is concluded by verifying that $\det(A) = \sigma^n$ and by using the fact that $P_e = 1 - P_c$. ■

Remark 4. *There are several alternative ways of expressing the probability of error in* (14) *such as*

$$\begin{aligned} 1 - P_e &= n! \frac{\mathrm{Vol}(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}{\sigma^n \mathrm{Vol}(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))} \\ &\overset{(a)}{=} \frac{1}{n!} \frac{\mathrm{Vol}(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}{\mathrm{Vol}\left(A(\mathcal{C}_{\mathcal{H}_\pi} \cap \mathcal{B}^{2n}(\mathbf{0}_{2n}, 1)) \right)} \\ &\overset{(b)}{=} \frac{\mathrm{Vol}(\mathcal{A} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}{\sigma^n \mathrm{Vol}(\mathcal{B}^{2n}(\mathbf{0}_{2n}, 1))}, \end{aligned}$$

where the equality in (a) *follows since* $\mathrm{Vol}(A\mathcal{S}) = |\det(A)| \mathrm{Vol}(\mathcal{S})$ *for any invertible matrix A and any set $\mathcal{S}$, and by computing the volume of the intersection of a ball and a cone $\mathcal{C}_{\mathcal{H}_\pi}$ and the equality in* (b) *follows by letting $\mathcal{A} = \cup_{\pi \in \mathcal{P}} \mathcal{C}_{\mathcal{H}_\pi}$, i.e., $\mathcal{A}$ is the collection of events of correct detection.*

B. Behavior of P_e as $n \to \infty$.

Using the expression for the probability of error in (14) together with a *covering* argument, we can find the first order logarithmic behavior of the probably of error, as stated in the next proposition.

Proposition 1. *Assume that $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ with $K_{\mathbf{N}} = \sigma^2 I_n$. Then, the probability of correctness can be upper and lower bounded as*

$$\frac{1}{n!} \le P_c \le \frac{1}{n!}\frac{\|A\|^{2n}}{\sigma^n}, \tag{19a}$$

where $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & \sigma I_n \end{bmatrix} \in \mathbb{R}^{2n}$ *and*

$$\|A\| = \left(\frac{(\sigma^4+4)^{\frac{1}{2}}}{2} + \frac{\sigma^2}{2} + 1 \right)^{\frac{1}{2}}. \tag{19b}$$

Consequently,

$$\lim_{n\to\infty} \frac{\log \frac{1}{P_c}}{\log(n!)} = 1. \tag{19c}$$

Proof: We start by deriving the lower bound on P_c,

$$\begin{aligned} P_c &= \sum_{\pi\in\mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{X}+\mathbf{N})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &\ge \sum_{\pi\in\mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{N})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi \right) \\ &= n! \frac{1}{(n!)^2}, \end{aligned}$$

where the inequality follows from the fact that if $(\mathbf{X}, \mathbf{N})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi$ then $(\mathbf{X}, \mathbf{X}+\mathbf{N})^T \in \mathcal{H}_\pi \times \mathcal{H}_\pi$.

Now to show the upper bound we use Theorem 2. We have

$$\begin{aligned} P_c &= n! \frac{\mathrm{Vol}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}(\mathbf{0}_{2n},1)\right)}{\sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},1)\right)} \\ &\overset{(a)}{\le} n! \frac{\mathrm{Vol}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap \mathcal{B}^{2n}(\mathbf{0}_{2n},\|A\|)\right)}{\sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},1)\right)} \\ &\overset{(b)}{=} \frac{\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},\|A\|)\right)}{n!\sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},1)\right)} \\ &\overset{(c)}{=} \frac{\|A\|^{2n}}{n!\sigma^n}, \end{aligned} \tag{20}$$

where the labeled (in)equalities follow from: (a) using the fact that $A\mathcal{B}^{2n}(\mathbf{0}_{2n},1) \subseteq \mathcal{B}^{2n}(\mathbf{0}_{2n},\|A\|)$; (b) computing the volume of the intersection of a ball and a cone $\mathcal{C}_{\mathcal{H}_\pi}$; and (c) using the fact that $\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},\|A\|)\right) = \|A\|^{2n}\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n},1)\right)$. The proof is concluded by using the fact that the spectral norm of A is given by the largest singular value of A, that is given by (19b). ■

Remark 5. *The fact that $P_e = 0$ at $\sigma = 0$, together with the upper bound in* (19a)*, suggests that to achieve a small probability of error for even moderate values of n the value of σ must be extremely small.*

Remark 6. *The upper bound on P_c in* (19a) *is a function of $\|A\|^{2n}/\sigma^n$. It is not difficult to see that, for $\sigma > 0$, we have the following range of parameters for $\|A\|^2/\sigma$*

$$\frac{\|A\|^2}{\sigma} \in \left[\sqrt{2}+1, \infty\right), \tag{21}$$

where the minimum is achieved at $\sigma = \sqrt{2}$.

The convergence of the upper bound on P_c in (19a) for finite values of n is evaluated in Fig. 3. From Fig. 3 we observe that, since $\frac{\|A\|^2}{\sigma}$ is convex in $\sigma \in (0,\infty]$ with minimum value achieved at $\sigma = \sqrt{2}$ (see Remark 6), then the convergence of the bound to 1 in (19c) for $\sigma = \sqrt{2}$ is fastest among all other values of σ. From Fig. 3, we also note that: (i) for $\sigma \in (0, \sqrt{2}]$ the convergence of the bound to 1 is faster for higher values of σ, whereas (ii) for $\sigma \in [\sqrt{2}, \infty)$ the convergence of the bound to 1 is faster for smaller values of σ.

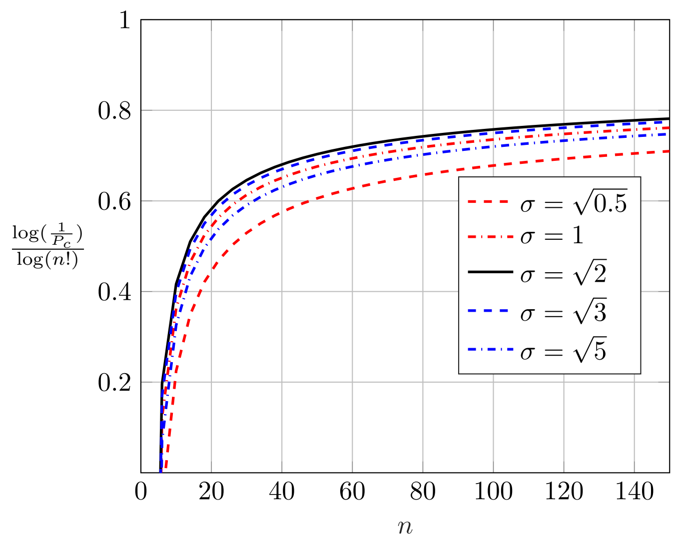

Fig. 3. Convergence of the upper bound on P_c in (19a) for finite values of n with $\sigma = \left\{\sqrt{0.5}, 1, \sqrt{2}, \sqrt{3}, \sqrt{5}\right\}$.

V. Conclusion

This paper has considered a problem of recovering/detecting a permutation of a sequence from an observation corrupted by Gaussian noise. The structure of the optimal decision regions was characterized for the case of i.i.d. noise and shown to be independent of the variance of the noise. Then, using the structure of the optimal decision regions, the probability of error was derived in terms of a volume of a region that consists of the intersection of a cone with a linear transformation of the unit radius $2n$-dimensional ball. Finally, this characterization of the error probability was used to determine its logarithmic behavior as the length of the sequence goes to infinity.

An interesting future direction, which is a subject of the current investigation, is to extend the result to colored Gaussian noise. Another interesting direction is to characterize the asymptotic behavior of the probability of error in the regimes of large and small noise variances.

REFERENCES

[1] 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.

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

[3] 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.

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

[5] M. Slawski and E. Ben-David, "Linear regression with sparsely permuted data," *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 1 –36, 2019.

[6] 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.

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

[8] D. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, 2017, pp. 1530–1539.

[9] 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, 2019.

[10] R. Ma, T. T. Cai, and H. Li, "Optimal permutation recovery in permuted monotone matrix model," *arXiv:1911.10604*, November 2019.

[11] S. Marano and P. K. Willett, "Algorithms and fundamental limits for unlabeled detection using types," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2022–2035, April 2019.

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