

Retrieving Data Permutations from Noisy Observations: High and Low Noise Asymptotics

Minoh Jeong[†], Alex Dytso^{*}, Martina Cardone[†]

[†] 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 permutation of an n -dimensional random vector $\mathbf{X}$ observed in Gaussian noise. First, a general expression for the probability of error is derived when a linear decoder (i.e., linear estimator followed by a sorting operation) is used. The derived expression holds with minimal assumptions on the distribution of $\mathbf{X}$ and when the noise has memory. Second, for the case of isotropic noise (i.e., noise with a diagonal scalar covariance matrix), the rates of convergence of the probability of error are characterized in the high and low noise regimes. In the low noise regime, for every dimension n , the probability of error is shown to behave proportionally to σ , where σ is the noise standard deviation. Moreover, the slope is computed exactly for several distributions and it is shown to behave quadratically in n . In the high noise regime, for every dimension n , the probability of correctness is shown to behave as $1/\sigma$, and the exact expression for the rate of convergence is also provided.

I. INTRODUCTION

The problem of recovering data permutations from noisy observations is becoming a common task of modern communication and computing systems. For example, systems based on data sorting operations, such as a recommender system or a data analysis system, make use of the data permutations and leverage the information that can be obtained from the data ordering. In particular, recommender systems clearly utilize the sorting information in order to optimize their next recommendation. As for the case of a recommender system, data analysis systems are also often interested in rankings of massive data sets rather than in the exact values of the data. In such systems, users may desire to enclose their data when it contains sensitive information. A common solution to privatize individual data is to add a sufficient amount of random noise to guarantee the desired privacy level [1]. However, adding too much noise can render the task of recovering a permutation impossible as the data will be too noisy. Therefore, for a given noise level, it is important to understand the fundamental limits of the recovering data permutations problem.

In this work, following preliminary works in [2] and [3], we study the problem of permutation recovery in the framework of an M -ary hypothesis testing. The specific goal of this paper is to study fundamental limits of such problem under the constraint that a linear decoder (i.e., linear estimator followed by a sorting operation) is employed. Studying linear decoders is interesting for several reasons. First, as it was shown in [2]

linear decoders are optimal (i.e., they lead to the smallest probability of error) when the noise is isotropic, and the distribution of the input data is exchangeable. Second, the optimal decoder can be linear even if the noise is colored; see [3] for the exact conditions. Third, linear decoders have at most polynomial complexity in the data dimension and hence, they are suitable for practical implementations.

The structure of the paper is as follows. In Section II, we introduce the notation and we formally define the problem. In Section III, we characterize the probability of error when linear decoders are used. The derived expression holds with minimal assumptions on the distribution of the data and holds when the noise has memory. In Section IV, we utilize the expression for the error probability derived in Section III and we characterize the *asymptotic* behavior of the probability of error for the isotropic noise case (i.e., when the noise covariance matrix is a diagonal scalar matrix) in the low and high noise regimes. For example, we show that the probability of error linearly increases in σ (i.e., the standard deviation of the noise) in the low noise regime (i.e., when $\sigma \rightarrow 0$). We derive the exact slope and we show it to be at most a quadratic function of the data dimension for a general class of distributions. In addition, we show that the behavior of the probability of correctness in the high noise regime (i.e., when $\sigma \rightarrow \infty$) is proportional to $1/\sigma$, and we characterize the exact slope.

A. Related work

Permutation associated estimation problems have recently gained significant importance and are studied in various fields [4]–[18]. The ranking (e.g., data permutation) estimation problem under a joint Gaussian distribution was investigated in [4]–[7]. In particular, in [4] the author considered a pairwise ordering for the bivariate case; the extended version to the n -dimension was considered in [5]. The generalization of the assumption of a Gaussian distribution to an elliptically contoured distribution can be found in [6], [7]. The authors in [4]–[7] analyzed the structure of the covariance matrix that maximizes the probability of correctness of such estimation problems using the minimum mean square error (MMSE) estimator. In [3], the MMSE estimator was shown to be the only linear estimator that achieves the minimum probability of error for the ranking estimation problem. Most of recent works study a problem based on a linear regression framework premultiplied by an unknown permutation matrix, which suitably models the problem with unknown labels. In [8], the

feature matching problem in computer vision was formulated as a permutation recovery problem. The multivariate linear regression model with an unknown permutation was studied in [9], [10]. The authors provided necessary and sufficient conditions on the signal-to-noise ratio for an exact permutation recovery and characterized the minimax prediction error. The isotonic regression without data labels, namely the *uncoupled isotonic regression*, was discussed in [11]. Data estimation given randomly selected measurements – referred to as *unlabeled sensing* – was studied in [12]–[14]. In [12], the authors characterized a necessary condition on the dimension of the observation vector for uniquely recovering the original data in the noiseless case. A generalized framework of unlabeled sensing was presented in [15]–[17]. The estimation of a sorted vector based on noisy observations was proposed in [18], where the MMSE estimator on sorted data was characterized as a linear combination of estimators on the unsorted data.

II. NOTATION AND FRAMEWORK

Notation. Boldface upper case letters $\mathbf{X}$ denote vector random variables; the boldface lower case letter $\mathbf{x}$ indicates a specific realization of $\mathbf{X}$; $X_{i:n}$ denotes the i -th order statistics of $\mathbf{X}$; $\|\mathbf{X}\|$ is the norm 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 $\mathcal{A}$; for $\mathcal{A}$ 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. For a set $\mathcal{S} \subseteq \mathbb{R}^n$, $\text{Vol}(\mathcal{S})$ denotes the volume, i.e., the n -dimensional Lebesgue measure. For two n -dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, if for all $i \in [1 : n]$, the i -th element of $\mathbf{x}$ is larger than or equal to the i -th element of $\mathbf{y}$, then we use $\mathbf{x} \geq \mathbf{y}$. Finally, the multiplication of a matrix A by a set $\mathcal{B}$ is denoted and defined as $A\mathcal{B} = \{A\mathbf{x} : \mathbf{x} \in \mathcal{B}\}$.

We consider the framework in Fig. 1, where an n -dimensional random vector $\mathbf{X} \in \mathbb{R}^n$ is first generated according to a certain distribution and then 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 $K_{\mathbf{N}}$ is 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 studying the *probability of error* of the “data permutation recovery” problem formulated in [2], [3] that, given the observation of $\mathbf{Y}$, seeks to retrieve the permutation (among the $n!$ possible ones) according to which the vector $\mathbf{X}$ is sorted. Specifically, this problem can be formulated within a hypothesis testing framework with $n!$ hypotheses $\mathcal{H}_{\pi}, \pi \in \mathcal{P}$, where $\mathcal{P}$ is the collection of all permutations of the elements of $[1 : n]$, and where $\mathcal{H}_{\pi}$ is the hypothesis that $\mathbf{X}$ is an n -dimensional vector sorted according to the permutation $\pi \in \mathcal{P}$, that is

$$\mathcal{H}_{\pi} = \{\mathbf{x} \in \mathbb{R}^n : x_{\pi_1} \leq x_{\pi_2} \leq \dots \leq x_{\pi_n}\}, \quad (1)$$

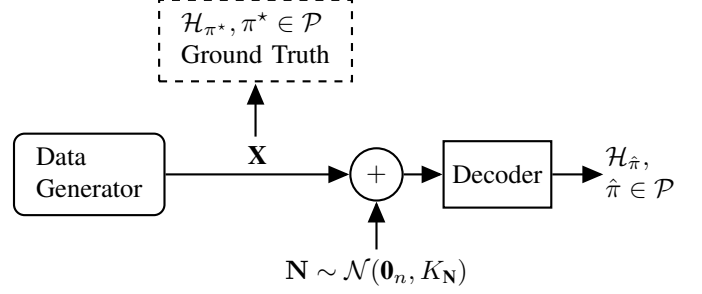


Fig. 1. Graphical representation of the considered framework.

with $x_{\pi_i}, i \in [1 : n]$ being the π_i -th element of $\mathbf{x}$, and $\pi_i, i \in [1 : n]$ being the i -th element of π . Given this, the *optimal* decoder in Fig. 1 will output $\mathcal{H}_{\hat{\pi}}, \hat{\pi} \in \mathcal{P}$ such that

$$\mathcal{H}_{\hat{\pi}} : \hat{\pi} = \underset{\pi \in \mathcal{P}}{\operatorname{argmin}} \{\Pr(\mathcal{H}_{\pi} \neq \mathcal{H}_{\pi^*})\}, \quad (2)$$

where π^* denotes the permutation according to which the random vector $\mathbf{X}$ is sorted. In particular, the decoder will declare that the input vector $\mathbf{x} \in \mathcal{H}_{\pi}$ if and only if the observation vector $\mathbf{y} \in \mathcal{R}_{\pi, K_{\mathbf{N}}}$, where $\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{P}$ are the so-called *optimal* decision regions¹, which can be derived by leveraging the maximum a posteriori probability (MAP) criterion [19, Appendix 3C] and are given by [2], [3]

$$\mathcal{R}_{\pi, K_{\mathbf{N}}} = \left\{ \mathbf{y} \in \mathbb{R}^n : f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\pi}) > \max_{\substack{\tau \in \mathcal{P} \\ \tau \neq \pi}} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\tau}) \right\}, \quad (3)$$

where $f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\pi}) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_{\pi})\Pr(\mathcal{H}_{\pi})$ with $f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_{\pi})$ denoting the conditional probability density function of $\mathbf{Y}$ given that $\mathbf{X} \in \mathcal{H}_{\pi}$. 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 $\mathbf{y} \in \{\mathcal{R}_{\pi, K_{\mathbf{N}}}, \pi \in \mathcal{S}, \mathcal{S} \subseteq \mathcal{P}, |\mathcal{S}| > 1\}$, then one of the hypotheses $\mathcal{H}_{\pi}, \pi \in \mathcal{S}$ is arbitrarily selected.

III. PROBABILITY OF ERROR WITH LINEAR DECODER

In this section, we focus on characterizing the probability of error of the data permutation recovery problem introduced in Section III. Given the hypothesis and decision regions defined in (1) and (3), we have that the error probability P_e is given by

$$P_e = 1 - P_c, \quad (4a)$$

$$P_c = \sum_{\pi \in \mathcal{P}} \Pr(\{\mathbf{Y} \in \mathcal{R}_{\pi, K_{\mathbf{N}}}\} \cap \{\mathbf{X} \in \mathcal{H}_{\pi}\}), \quad (4b)$$

where P_c is the probability of correctness.

In particular, we assess the probability of error when a *linear* decoder is employed. This decoder first computes a permutation-independent linear transformation $\mathbf{y}_{\ell}$ of $\mathbf{y}$, i.e., $\mathbf{y}_{\ell} = A\mathbf{y} + \mathbf{b}$, where $A \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$ are the same for all permutations, and then it outputs the permutation according

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

to which $\mathbf{y}_\ell$ is sorted. The decision regions in (3) when a linear decoder is used become

$$\mathcal{R}_{\pi, K_N} = A\mathcal{H}_\pi + \mathbf{b}, \quad \forall \pi \in \mathcal{P}. \quad (5)$$

Our choice of assessing the probability of error performance of a linear decoder stems primarily from its low complexity (at most polynomial in n) compared to a brute force evaluation of the optimal test (3), which has a practically prohibitive complexity of $n!$. Moreover, for the case $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ it has been shown in [3] that a linear decoder is indeed optimal, i.e., it minimizes the probability of error, under certain conditions on the noise covariance matrix K_N .

We next derive an expression for the probability of error when a linear decoder is used. Towards this end, for each $\pi \in \mathcal{P}$, we define a matrix $T_\pi \in \mathbb{R}^{(n-1) \times n}$ such that

$$(T_\pi)_{i,j} = 1_{\{j=\pi_{i+1}\}} - 1_{\{j=\pi_i\}}, \quad (6)$$

where $1_{\{x=y\}} = 1$ if and only if $x = y$ and is equal to zero otherwise. For instance, let $n = 4$ and consider $\pi = \{4, 2, 1, 3\}$; then, we have that

$$T_{\{4,2,1,3\}} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}.$$

The theorem below provides an expression for the error probability of the data permutation recovery problem when a linear decoder is used.

Theorem 1. *Let $\mathbf{X}$ be an exchangeable random vector². Then, for any invertible A and $\mathbf{b}$ defined in (5) and any noise covariance matrix K_N , the probability of error is given by*

$$P_e = 1 - \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \mathbb{E} [Q_{\tilde{K}_\pi}(-T_\pi A^{-1}(\mathbf{X} - \mathbf{b})) \mid \mathbf{X} \in \mathcal{H}_\pi], \quad (7)$$

where $\tilde{K}_\pi = T_\pi A^{-1} K_N A^{-T} T_\pi^T \in \mathbb{R}^{(n-1) \times (n-1)}$ with $T_\pi, \pi \in \mathcal{P}$ given by (6), and where $Q_{\tilde{K}_\pi}(\cdot)$ is the multivariate Gaussian Q-function with covariance $\tilde{K}_\pi$.

Proof: By substituting the decision regions in (5) inside (4) and by using the Bayes' theorem, we obtain

$$\begin{aligned} P_e &= \sum_{\pi \in \mathcal{P}} \Pr(\mathbf{Y} \in A\mathcal{H}_\pi + \mathbf{b} \mid \mathbf{X} \in \mathcal{H}_\pi) \Pr(\mathbf{X} \in \mathcal{H}_\pi) \\ &\stackrel{(a)}{=} \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \Pr(\mathbf{X} + K_N^{\frac{1}{2}} \mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\pi \mid \mathbf{X} \in \mathcal{H}_\pi) \\ &\stackrel{(b)}{=} \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \mathbb{E} \left[\Pr(\mathbf{X} + K_N^{\frac{1}{2}} \mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\pi \mid \mathbf{X}) \mid \mathbf{X} \in \mathcal{H}_\pi \right], \quad (8) \end{aligned}$$

where (a) follows from the fact that $\mathbf{X}$ is exchangeable and hence, $\Pr(\mathbf{X} \in \mathcal{H}_\pi) = \frac{1}{n!}$, $\forall \pi \in \mathcal{P}$ and letting $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$, and (b) is due to the law of total expectation.

²A sequence of random variables $U_1, \dots, U_n$ is said to be exchangeable if, for any permutation $(\pi_1, \dots, \pi_n)$ of the indices $[1 : n]$, we have that $(U_1, \dots, U_n)$ is equal in distribution to $(U_{\pi_1}, \dots, U_{\pi_n})$.

We now focus on the conditional probability inside the conditional expectation in (8). For each $\mathcal{H}_\pi$, $\forall \pi \in \mathcal{P}$ we have

$$\begin{aligned} &\Pr(\mathbf{X} + K_N^{\frac{1}{2}} \mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\pi \mid \mathbf{X}) \\ &= \Pr(A^{-1}(\mathbf{X} - \mathbf{b}) + A^{-1} K_N^{\frac{1}{2}} \mathbf{Z} \in \mathcal{H}_\pi \mid \mathbf{X}) \\ &= \Pr(A^{-1}(\mathbf{X} - \mathbf{b}) + \mathbf{U} \in \mathcal{H}_\pi \mid \mathbf{X}), \quad (9) \end{aligned}$$

where the last equality follows by letting $\mathbf{U} = A^{-1} K_N^{\frac{1}{2}} \mathbf{Z}$. Note that $\mathbf{U} \sim \mathcal{N}(\mathbf{0}_n, A^{-1} K_N A^{-T})$.

Then, given $\mathbf{X}$, the event inside the conditional probability in (9) can be expressed as

$$\begin{aligned} &\{A^{-1}(\mathbf{X} - \mathbf{b}) + \mathbf{U} \in \mathcal{H}_\pi\} \\ &= \bigcap_{k=1}^{n-1} \left\{ (A^{-1}(\mathbf{X} - \mathbf{b}))_{\pi_k} - (A^{-1}(\mathbf{X} - \mathbf{b}))_{\pi_{k+1}} \leq U_{\pi_{k+1}} - U_{\pi_k} \right\} \\ &= \{-T_\pi A^{-1}(\mathbf{X} - \mathbf{b}) \leq T_\pi \mathbf{U}\}, \quad (10) \end{aligned}$$

where the last equality follows by using the definition of $T_\pi, \pi \in \mathcal{P}$ in (6). By introducing a random vector $\mathbf{V}_\pi = T_\pi \mathbf{U} \sim \mathcal{N}(\mathbf{0}_{n-1}, \tilde{K}_\pi)$, where $\tilde{K}_\pi = T_\pi A^{-1} K_N A^{-T} T_\pi^T$, we have an equivalent expression for (9) as $\Pr(\mathbf{X} + K_N^{\frac{1}{2}} \mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\pi \mid \mathbf{X}) = \Pr(-T_\pi A^{-1}(\mathbf{X} - \mathbf{b}) \leq \mathbf{V}_\pi \mid \mathbf{X})$. By substituting this inside (8), we obtain

$$\begin{aligned} P_e &= \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \mathbb{E} [\Pr(-T_\pi A^{-1}(\mathbf{X} - \mathbf{b}) \leq \mathbf{V}_\pi \mid \mathbf{X}) \mid \mathbf{X} \in \mathcal{H}_\pi] \\ &= \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \mathbb{E} [Q_{\tilde{K}_\pi}(-T_\pi A^{-1}(\mathbf{X} - \mathbf{b})) \mid \mathbf{X} \in \mathcal{H}_\pi], \quad (11) \end{aligned}$$

where the last equality follows by letting $Q_{\tilde{K}_\pi}(\cdot)$ be the multivariate Gaussian Q-function with covariance $\tilde{K}_\pi$. We conclude the proof of Theorem 1 by using $P_e = 1 - P_c$. ■

We highlight that (7) holds with minimal assumption on the distribution of $\mathbf{X}$ (i.e., exchangeability) and hence, it can be used to study the error probability of the data permutation recovery problem in various noise settings, e.g., noise has memory, noise is isotropic. In the remaining of this paper, we will focus on the isotropic noise scenario, i.e., we assume that K_N is a diagonal scalar matrix.

IV. ISOTROPIC NOISE

We here study the error probability of the data permutation recovery problem when the noise is isotropic, i.e., $K_N = \sigma^2 I_n$. Under this assumption, the regions $\mathcal{R}_{\pi, K_N}, \pi \in \mathcal{P}$ in (5) depends on K_N only through the parameter σ and hence, we let $\mathcal{R}_{\pi, K_N} = \mathcal{R}_{\pi, \sigma}$. Moreover, when $\mathbf{X}$ is exchangeable, it has been shown in [2] that $\mathcal{R}_{\pi, \sigma} = \mathcal{H}_\pi, \pi \in \mathcal{P}$, i.e., for the isotropic noise setting the optimal decoder is indeed linear and hence, the probability of error in Theorem 1 is the minimum.

In Section IV-A, we will evaluate the probability of error P_e in (7) when $K_N = \sigma^2 I_n$ and then in Section IV-B and Section IV-C, we will use this expression to derive the rates of convergence of P_e in the low noise regime (i.e., $\sigma \rightarrow 0$) and high noise regime (i.e., $\sigma \rightarrow \infty$), respectively.

A. Probability of Error

Under the assumption $K_{\mathbf{N}} = \sigma^2 I_n$ we have that $\mathcal{R}_{\pi, \sigma} = \mathcal{H}_{\pi}, \pi \in \mathcal{P}$ [2] and hence, with reference to [5], we have that $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$. Moreover, by substituting these values inside $\tilde{K}_{\pi} \in \mathbb{R}^{(n-1) \times (n-1)}$ in Theorem 1, we obtain

$$\tilde{K}_{\pi} = T_{\pi} A^{-1} K_{\mathbf{N}} A^{-T} T_{\pi}^T = \sigma^2 T_{\pi} T_{\pi}^T = \sigma^2 \tilde{K}, \quad (12)$$

$$(\tilde{K})_{i,j} = \begin{cases} 2 & i = j, \\ -1 & i = j + 1 \text{ and } j = i + 1, \\ 0 & \text{otherwise,} \end{cases}$$

that is $\tilde{K} \in \mathbb{R}^{(n-1) \times (n-1)}$ is a tridiagonal Toeplitz matrix.

The probability of error in the isotropic noise scenario is then given by the next corollary.

Corollary 1. *Let $\mathbf{X}$ be an exchangeable random vector and let $K_{\mathbf{N}} = \sigma^2 I_n$. Then, for an arbitrary $\pi \in \mathcal{P}$, the probability of error is given by*

$$P_e = 1 - \mathbb{E}[Q_{\sigma^2 \tilde{K}}(-T_{\pi} \mathbf{X}) \mid \mathbf{X} \in \mathcal{H}_{\pi}], \quad (13)$$

where $\tilde{K}$ is defined in (12) and where $Q_{\sigma^2 \tilde{K}}(\cdot)$ is the multivariate Gaussian Q-function with covariance $\sigma^2 \tilde{K}$.

Proof: By substituting the expression of $\tilde{K}_{\pi}$ in (12) inside (7), we obtain

$$P_e = 1 - \frac{1}{n!} \sum_{\pi \in \mathcal{P}} \mathbb{E}[Q_{\sigma^2 \tilde{K}}(-T_{\pi} \mathbf{X}) \mid \mathbf{X} \in \mathcal{H}_{\pi}]. \quad (14)$$

We note that $\sigma^2 \tilde{K}$ and the distribution of $T_{\pi} \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_{\pi}$ are independent of $\pi \in \mathcal{P}$ and hence, the conditional expectation in (14) is constant in $\pi \in \mathcal{P}$. Since $|\mathcal{P}| = n!$, we obtain

$$P_e = 1 - \mathbb{E}[Q_{\sigma^2 \tilde{K}}(-T_{\tau} \mathbf{X}) \mid \mathbf{X} \in \mathcal{H}_{\tau}], \quad (15)$$

where $\tau \in \mathcal{P}$ can be arbitrary. ■

We note that (14) is a function of σ and hence, in what follows we will use $P_e(\sigma)$ to highlight this dependence.

B. Low Noise Asymptotic

We here focus on the asymptotic behavior of the probability of error in the low noise regime (i.e., $\sigma \rightarrow 0$). In particular, the next result, proved in [20, Appendix A], shows that the probability of error in this regime is approximately linear in σ .

Theorem 2. *Let $\mathbf{X}$ consist of n i.i.d. random variables generated according to X . Let X' be an independent copy of X and assume that*

$$f_{X-X'}(x) < \infty, \quad \forall x \in \mathbb{R}. \quad (16)$$

Then,

$$\lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}, \quad (17)$$

where $W_i = X_{i+1:n} - X_{i:n}$, $i \in [1 : n-1]$.

Remark 1. *The i.i.d. assumption on $\mathbf{X}$ in Theorem 2 can be relaxed to the case of exchangeable $\mathbf{X}$, provided that the following holds: for $1 \leq i < j \leq n-1$,*

$$f_{W_i}(u) < \infty, \quad \forall u \in \mathbb{R}_+, \quad (18a)$$

$$f_{W_i, W_j}(u, v) < \infty, \quad \forall (u, v) \in \mathbb{R}_+^2. \quad (18b)$$

Then, under these conditions, for the case of an exchangeable $\mathbf{X}$ we have the same result as in Theorem 2.

Remark 2. *The quantity $f_{W_i}(0^+)$ in (17) can be computed as follows [21]*

$$f_{W_i}(0^+) = \frac{n! \int_{-\infty}^{\infty} F(x)^{i-1} (1-F(x))^{n-i-1} f^2(x) dx}{(i-1)!(n-i-1)!} \\ = \frac{n! \mathbb{E}[U^{i-1} (1-U)^{n-i-1} f(F^{-1}(U))]}{(i-1)!(n-i-1)!}, \quad (19)$$

where the last step uses the probability integral transformation theorem and the quantile function theorem [22] with $U \sim \text{Unif}(0, 1)$.

We now show that, under the condition $\sup_{x \in \mathbb{R}} f_X(x) = c$, the asymptotic behavior of the probability of error in the low noise regime for an i.i.d. $\mathbf{X}$ is upper bounded by $O(n^2)$. In particular, we have the following lemma.

Lemma 1. *Assume that $\sup_{x \in \mathbb{R}} f_X(x) = c$, where $c \in \mathbb{R}$ is a constant. Then,*

$$\lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} \leq c \frac{n(n-1)}{\sqrt{\pi}}. \quad (20)$$

Proof: By using the expression in (19), we have that

$$f_{W_i}(0^+) = \frac{n! \mathbb{E}[U^{i-1} (1-U)^{n-i-1} f(F^{-1}(U))]}{(i-1)!(n-i-1)!} \\ \leq c \frac{n! \mathbb{E}[U^{i-1} (1-U)^{n-i-1}]}{(i-1)!(n-i-1)!} \\ = c \frac{n! \int_0^1 x^{i-1} (1-x)^{n-i-1} dx}{(i-1)!(n-i-1)!} \\ = c \frac{n!}{(i-1)!(n-i-1)!} \frac{\Gamma(n-i)\Gamma(i)}{\Gamma(n)} = cn,$$

where $\Gamma(\cdot)$ is the gamma function and where the inequality follows by using the bound $f(F^{-1}(U)) \leq c = \sup_{x \in \mathbb{R}} f(x)$. Hence, (17) can be upper bounded as

$$\lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}} \leq c \frac{(n-1)n}{\sqrt{\pi}}. \quad (21)$$

This concludes the proof of Lemma 1. ■

We conclude this section by providing some examples of (17) for a few distributions.

Example 1. Consider $X \sim \text{Unif}(a, b)$, $0 \leq a < b < \infty$. Then,

$$\lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} = \frac{n(n-1)}{(b-a)\sqrt{\pi}}. \quad (22)$$

The proof of (22) can be found in [20, Appendix D-A]. ◇

Example 2. Consider $X \sim \text{Exp}(\lambda)$, $\lambda > 0$. Then,

$$\lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} = \frac{\lambda n(n-1)}{2\sqrt{\pi}}. \quad (23)$$

The proof of (23) can be found in [20, Appendix D-B]. $\diamond$

Example 3. Consider $X \sim \mathcal{N}(0, 1)$. Then,

$$\frac{\sqrt{2}n(n-1)}{6\pi} \leq \lim_{\sigma \rightarrow 0} \frac{P_e(\sigma)}{\sigma} \leq \frac{n(n-1)}{\sqrt{2}\pi}. \quad (24)$$

Note that the upper bound in (24) follows from Lemma 1, where we used the fact that $c = \sup_{x \in \mathbb{R}} f(x) = \frac{1}{\sqrt{2\pi}}$. For the lower bound we use the following inequality [23, Lemma 10.1]:

$$f(F^{-1}(u)) \geq \sqrt{\frac{2}{\pi}} \min\{u, 1-u\} \geq \sqrt{\frac{2}{\pi}} u(1-u), \quad (25)$$

where the last step follows since $\min\{a, b\} \geq \frac{ab}{a+b}$ for all $a > 0, b > 0$. Combining the expression in (19) and the bound in (25), we arrive at the following lower bound,

$$f_{W_i}(0^+) \geq \sqrt{\frac{2}{\pi}} \frac{n! \mathbb{E}[U^i (1-U)^{n-i}]}{(i-1)!(n-i-1)!} = \sqrt{\frac{2}{\pi}} \frac{i(n-i)}{(n+1)}, \quad (26)$$

which implies the lower bound in (24). $\diamond$

C. High Noise Asymptotic

We now focus on the asymptotic behavior of the probability of error in the high noise regime (i.e., $\sigma \rightarrow \infty$). It is not difficult to argue that if $\mathbf{X}$ is exchangeable, then we have that

$$\lim_{\sigma \rightarrow \infty} P_e(\sigma) = 1 - \frac{1}{n!} = P_e(\infty). \quad (27)$$

The interpretation is that if σ is large, then the output $\mathbf{Y}$ carries no information of $\mathbf{X}$, and the decoder can only rely on the prior knowledge; hence, the best thing that the decoder can do is to guess one of the $n!$ hypotheses.

The next result, proved in [20, Appendix B], sharpens the limit in (27) by finding the rate of convergence.

Theorem 3. Let $\mathbf{X}$ be an exchangeable random vector such that $\mathbb{E}[\|\mathbf{X}\|] < \infty$. Then,

$$\lim_{\sigma \rightarrow \infty} \frac{P_e(\infty) - P_e(\sigma)}{\frac{1}{\sigma}} = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \alpha_i \mathbb{E}[W_i], \quad (28)$$

where $W_i = X_{i+1:n} - X_{i:n}$, $i \in [1 : n-1]$ and

$$\alpha_i = \frac{\text{Vol}(\mathcal{E}(\mathbf{0}_{n-1}, i) \cap \mathcal{H}_{[1:n-1]})}{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1))}, \quad (29)$$

where $\mathcal{H}_{[1:n-1]}$ is defined in (1), $\mathcal{B}(\mathbf{0}_{n-1}, 1)$ is the $(n-1)$ -dimensional ball centered at the origin with unitary radius, and $\mathcal{E}(\mathbf{0}_{n-1}, i)$ is the $(n-1)$ -dimensional ellipsoid centered at the origin with unit radii along standard axes except a $\frac{1}{\sqrt{2}}$ radius along the i -th axis.

Finding a closed-form expression for the α_i 's in (29) does not appear to be an easy task. In the next lemma, we provide upper and lower bounds on the α_i 's, which lead to expressions that are amenable to computations.

Lemma 2. In the high noise regime, the convergence rate of the probability of correctness can be bounded as

$$\frac{\mathbb{E}[R_n]}{\sqrt{\pi}(n-1)!2^{\frac{n}{2}}} \leq \lim_{\sigma \rightarrow \infty} \frac{P_e(\infty) - P_e(\sigma)}{\frac{1}{\sigma}} \leq \frac{\mathbb{E}[R_n]}{\sqrt{2\pi}(n-1)!},$$

where $R_n = X_{n:n} - X_{1:n}$.

Proof: We start by observing that

$$\mathcal{B}(\mathbf{0}_{n-1}, 2^{-\frac{1}{2}}) \subset \mathcal{E}(\mathbf{0}_{n-1}, i) \subset \mathcal{B}(\mathbf{0}_{n-1}, 1), \quad (30)$$

that is, the ellipsoid $\mathcal{E}(\mathbf{0}_{n-1}, i)$: (i) contains the ball $\mathcal{B}(\mathbf{0}_{n-1}, 2^{-\frac{1}{2}})$ since $\mathcal{E}(\mathbf{0}_{n-1}, i)$ has minimum radius equal to $2^{-\frac{1}{2}}$; and (ii) is contained inside the ball $\mathcal{B}(\mathbf{0}_{n-1}, 1)$ since $\mathcal{E}(\mathbf{0}_{n-1}, i)$ has maximum radius equal to 1.

Thus, from (30) we obtain

$$\alpha_i \leq \frac{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1) \cap \mathcal{H}_{[1:n-1]})}{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1))} = \frac{1}{(n-1)!}, \quad (31)$$

where the last equality follows since $\mathcal{H}_{[1:n-1]}$ is a cone that occupies a $\frac{1}{(n-1)!}$ portion of the space and hence, $\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1) \cap \mathcal{H}_{[1:n-1]}) = \frac{1}{(n-1)!} \text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1))$.

Similarly, from (30) we obtain

$$\begin{aligned} \alpha_i &\geq \frac{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 2^{-\frac{1}{2}}) \cap \mathcal{H}_{[1:n-1]})}{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1))} \\ &= \left| \det \left(2^{-\frac{1}{2}} I_{n-1} \right) \right| \frac{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1) \cap \mathcal{H}_{[1:n-1]})}{\text{Vol}(\mathcal{B}(\mathbf{0}_{n-1}, 1))} \\ &= \frac{1}{2^{\frac{n-1}{2}} (n-1)!}, \end{aligned} \quad (32)$$

where in the first equality we have used the facts that: (i) $2^{\frac{1}{2}} I_{n-1} \mathcal{B}(\mathbf{0}_{n-1}, 2^{-\frac{1}{2}}) = \mathcal{B}(\mathbf{0}_{n-1}, 1)$, (ii) $2^{\frac{1}{2}} I_{n-1} \mathcal{H}_{[1:n-1]} = \mathcal{H}_{[1:n-1]}$, and (iii) $\text{Vol}(AS) = |\det(A)| \text{Vol}(S)$ for any invertible matrix A and any set S .

The proof of Lemma 2 is concluded by substituting (31) and (32) into (28) and by using the fact that

$$\sum_{i=1}^{n-1} \mathbb{E}[W_i] = \mathbb{E}[R_n], \quad (33)$$

where $R_n = X_{n:n} - X_{1:n}$ denotes the range [24] of $\mathbf{X}$. $\blacksquare$

We conclude this section by providing some examples of the range R_n for a few common distributions (see [20, Appendix E] for the detailed computations). In particular, these examples show that the term $\frac{1}{(n-1)!}$ dominates in the expression of the rate for several distribution of interest.

Example 1. Consider $X \sim \text{Unif}(a, b)$, $0 \leq a < b < \infty$. Then,

$$\mathbb{E}[R_n] = (b-a)(n-1)(n+1)^{-1}.$$

Example 2. Consider $X \sim \text{Exp}(\lambda)$, $\lambda > 0$. Then,

$$\mathbb{E}[R_n] = \frac{1}{\lambda} \sum_{k=1}^{n-1} \frac{1}{k} = O\left(\frac{1}{\lambda} \log(n)\right). \quad (34)$$

Example 3. Let X be γ^2 -sub-Gaussian³. Then [23],

$$\mathbb{E}[R_n] \leq 2\sqrt{2\gamma^2 \log(n)}.$$

³A random variable X is γ^2 -sub-Gaussian if $\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leq e^{\lambda^2 \gamma^2}$ for all $\lambda \in \mathbb{R}$.

REFERENCES

- [1] C. Dwork, “Differential privacy: A survey of results,” in *International conference on theory and applications of models of computation*. Springer, 2008, pp. 1–19.
- [2] M. Jeong, A. Dytso, M. Cardone, and H. V. Poor, “Recovering structure of noisy data through hypothesis testing,” in *Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)*, June 2020, pp. 1307–1312.
- [3] —, “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.
- [4] S. R. Searle *et al.*, “Prediction, mixed models, and variance components,” 1973.
- [5] 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.
- [6] 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.
- [7] —, “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.
- [8] 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.
- [9] 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.
- [10] —, “Denoising linear models with permuted data,” in *Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 446–450.
- [11] 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.
- [12] 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.
- [13] S. Haghighatshoar and G. Caire, “Signal recovery from unlabeled samples,” *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1242–1257, March 2018.
- [14] 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.
- [15] I. Dokmanić, “Permutations unlabeled beyond sampling unknown,” *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 823–827, April 2019.
- [16] 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.
- [17] M. C. Tsakiris, “Eigenspace conditions for homomorphic sensing,” *arXiv:1812.07966*, April 2019.
- [18] 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.
- [19] S. M. Kay, *Fundamentals of Statistical Signal Processing, vol. 2: Detection Theory*. Prentice Hall PTR, 1998.
- [20] M. Jeong, A. Dytso, and M. Cardone, “Retrieving data permutations from noisy observations: High and low noise asymptotics,” *arXiv:2105.03015*, 2021.
- [21] 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>
- [22] J. E. Angus, “The probability integral transform and related results,” *SIAM review*, vol. 36, no. 4, pp. 652–654, 1994.
- [23] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- [24] H. A. David and H. N. Nagaraja, “Order statistics,” *Encyclopedia of statistical sciences*, 2004.