Data Permutation Recovery from Noisy Data: Error Probability and Privacy

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Minoh Jeong

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Prof. Martina Cardone

May, 2024

Acknowledgements

I am immensely grateful to my supervisor, Professor Martina Cardone, for her unwavering support, guidance, and encouragement throughout this research endeavor. Her expertise, mentorship, and dedication have been invaluable in shaping the trajectory of this thesis and my PhD journey. I am deeply appreciative of the time she invested in providing constructive feedback and insightful suggestions that significantly enhanced the quality of this work.

I extend my sincere gratitude to Dr. Alex Dytso for his collaboration and contributions to this research project. His expertise in statistics and mathematical analysis has broadened the scope of my investigation and enriched the depth of my analysis skills. I am grateful for his willingness to share his knowledge and engage in fruitful discussions that have propelled this research forward.

I would like to express my heartfelt thanks to the members of my thesis committee, Professor Soheil Mohajer and Professor Mingyi Hong, for their invaluable guidance, feedback, and support throughout this academic journey. From the preliminary written and oral exams to final defense, their expertise and constructive criticism have played a pivotal role in refining the ideas presented in this thesis. I am grateful for the time they dedicated to reviewing my work and offering invaluable insights that have contributed to its overall quality.

Special thanks go to my colleagues Sarthak, Jaimin, Mohammad, Hamidreza in our group for their camaraderie and for valuable discussion. I am grateful for the numerous discussions, shared knowledge, and the collective effort that has made my research experience enjoyable and fulfilling. I would also thank to all my friends in our departments, in Korean EECS graduate student group, and in my home country.

I am deeply thankful to my family for their unconditional love and support. To my parents, who have always believed in me and provided the foundation for my education, and to my siblings, who have been my constant source of inspiration and motivation.

Dedication

This thesis is dedicated to my parents, Suhyeon and Woohyeon. Your love, sacrifices, and encouragement have been the cornerstone of my success. From the earliest days of my education, you have been my biggest supporters, instilling in me a love for learning and a drive to achieve. Your belief in me has provided the strength and motivation to persevere through every challenge. Without your constant support and guidance, this achievement would not have been possible. Every milestone I reach is a testament to your dedication and love.

To my siblings, Minsoo and Soyeong, your constant encouragement and unwavering belief in my potential have been a source of inspiration and strength. You have always been there to celebrate my successes and provide comfort during difficult times. Your love, support, and occasional tough love have shaped me into who I am today. Your belief in me has been a guiding light, helping me navigate the complexities of this journey.

To my girl friend, Min, thank you for your endless patience, understanding, and love. Your support has been my anchor, keeping me grounded and motivated. You have been my beloved, best friend, and supporter. Your sacrifices, late nights of encouragement, and unwavering belief in my dreams have made this possible. Your celebration of my PhD defense and graduation made me glad, and it is an unforgettable time for me. Your support has been invaluable, and I am incredibly grateful for having you in my life.

To my friends, Minki, Junkyu, Yoshitaka, Sanghyun, Hunmin, Zaemyung, Taejun, Jeongseok, Junno and everyone, who have been a part of this incredible journey, I dedicate this work to you, and I could enjoy my Minnesota life because of you. Also, I would thank my friends in Korea, Wontae, Uikyun, Yooho, Moongyu, Ganghee, Jangho, Sungjun, Jaecheol, Hochan, Eungyo, and everyone, who have always welcomed me whenever I visited Korea.

List of Papers

Papers included in this thesis:

[1] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering structure of noisy data through hypothesis testing. In *2020 IEEE International Symposium on Information Theory (ISIT),* pages 1307–1312, 2020.

[2] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering data permutations from noisy observations: The linear regime. *IEEE Journal on Selected Areas in Information Theory,* 1(3):854–869, 2020.

[3] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: High and low noise asymptotics. In *2021 IEEE International Symposium on Information Theory (ISIT),* pages 1100–1105, 2021.

[4] Minoh Jeong, Alex Dytso, and Martina Cardone. Ranking recovery under privacy considerations. *Transactions on Machine Learning Research,* 2022.

[5] Minoh Jeong, Martina Cardone, and Alex Dytso. On the ranking recovery from noisy observations up to a distortion. In *2022 IEEE International Symposium on Information Theory (ISIT),* pages 1993–1998, 2022.

[6] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: Asymptotics. *IEEE Transactions on Information Theory,* 70(4):2999 –3017, 2024.

Other papers by the author not included in this thesis:

[7] Minoh Jeong, Alex Dytso, and Martina Cardone. Gradient of error probability of M-ary hypothesis testing problems under multivariate Gaussian noise. *IEEE Signal Processing Letters,* 27:1909–1913, 2020.

[8] Minoh Jeong, Alex Dytso, and Martina Cardone. Functional properties of the Ziv-Zakai bound with arbitrary inputs. In *2023 IEEE International Symposium on Information Theory (ISIT),* pages 2087–2092, 2023.

[9] Minoh Jeong, Alex Dytso, and Martina Cardone. A comprehensive study on Ziv-Zakai lower bounds on the MMSE. *arXiv preprint arXiv:2404.04366,* 2024. (submitted to *IEEE Transactions on Information Theory*).

[10] Minoh Jeong, Martina Cardone, and Alex Dytso. Demystifying the optimal performance of multi-class classification. In *Advances in Neural Information Processing Systems,* volume 36, pages 31638–31664, 2023.

[11] Minoh Jeong, Martina Cardone, and Alex Dytso. Data-driven estimation of the false positive rate of the Bayes binary classifier via soft labels. arXiv preprint arXiv:2401.15500, 2024. (accepted to *2024 IEEE International Symposium on Information Theory*).

[12] Mohammad Milanian, Minoh Jeong, and Martina Cardone. On the secrecy capacity of 1-2-1 atomic networks. arXiv preprint arXiv:2405.05823, 2024. (accepted to *2024 IEEE International Symposium on Information Theory*).

[13] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Characterization of the quality factor in spiral coil designs for high-frequency wireless power transfer systems using machine learning. In *2022 IEEE 23rd Workshop on Control and Modeling for Power Electronics (COMPEL),* pages 1–8, 2022.

[14] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Optimization of spiral coil design for WPT systems using machine learning. In *2023 IEEE Applied Power Electronics Conference and Exposition (APEC),* pages 822–828, 2023.

[15] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Design of a spiral coil

for high-frequency wireless power transfer systems using machine learning. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics,* 5(1):193–202, 2024.

Abstract

This doctoral thesis investigates the data permutation recovery problem from noisy observations, a fundamental challenge at the intersection of data science and privacy-preserving technologies. The main question is: *Given a noisy observation of data, according to which permutation was the original data sorted?* This thesis addresses both the exact and approximate permutation recovery problems under various noise conditions. It begins by formulating the permutation recovery problem within the statistical hypothesis testing framework, and characterizes the linear regime, where the true permutation can be optimally estimated by only a linear transformation of the observation and a sorting operation. In particular, under Gaussian data and noise, this thesis derives the necessary and sufficient conditions for the linear regime in terms of the noise covariance matrix. Subsequently, this thesis shifts the focus to the examination of the error probability associated with linear decoders in the presence of Gaussian noise, but arbitrary data distribution. This analysis reveals the noise-dominated nature of the permutation recovery problem, illustrating how the error probability scales in both low- and high-noise regimes, and providing insights into the behavior of the error probability under different noise and data distributions. Specifically, the error probability of the permutation recovery problem grows linearly in the noise standard deviation σ in the low-noise regime, implying the noise-dominated nature of the problem. Advancing into the realm of data privacy, the thesis explores the private ranking recovery problem, aiming to establish fundamental trade-offs between estimation accuracy and privacy. Leveraging differential privacy metrics, it evaluates the effectiveness of various privacy-preserving mechanisms while ensuring the fidelity of the ranking recovery. Regarding the generalized permutation recovery problem, this thesis also proposes an approximate version of it. It demonstrates that this approximate version leads to an error probability having sub-linear behavior in σ in the low-noise regime, which is a notable difference with respect to the exact recovery. This finding highlights the potential of approximate recovery in enhancing the robustness and efficiency of permutation recovery. In summary, the thesis contributes significantly to our understanding of permutation recovery in noisy and privacy-sensitive environments. The thesis not only advances theoretical foundations but also provides practical strategies for tackling the challenges inherent in recovering data permutations, thereby offering valuable perspectives for advancements in data processing and privacy-preserving techniques.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the evolving landscape of digital communications and data analytics, the challenge of accurately recovering original data permutations from noisy observations has emerged as a cornerstone challenge. For example, in the data analytics realm, recommender systems are often more interested in recovering the relative ranking of data points rather than the values of the data itself. Furthermore, users may desire to privatize their data before it is collected by an external party, and one suitable solution to privatize data is perturbing it with noise. Upon receiving the perturbed/noisy data, the recommender system will then need to recover the data permutation (e.g., ranking of users' interests) in order to provide the next recommendation. This poses the following question: *Given a noisy observation of a data set, according to which permutation was the original data sorted?* This problem encapsulates a fundamental challenge in the fields of data science: how to recover the relative rankings of data points, an aspect often more valuable than the data values themselves, in the face of privacy concerns and the inevitable presence of noise. The introduction of noise, whether as a byproduct of data collection or as a deliberate mechanism for ensuring data privacy, necessitates the development of sophisticated algorithms capable of estimating the original data permutation.

In an attempt to answer the question, this thesis investigates the intricate realm of permutation recovery in perturbed data systems. In particular, we formulate the noisy data permutation recovery problem, which consists of recovering the permutation of an original data vector of size n that has been perturbed by additive noise. We consider a Bayesian framework where data is generated according to a certain distribution, and the perturbation consists of adding noise.

In Chapter 2, preliminaries of the permutation recovery problem are provided. Specifically,

we summarize the notation utilized in this thesis, and formally describe the noisy data permutation recovery problem, which consists of recovering the permutation of an original data vector of size n that has been perturbed by noise. Using a statistical M-ary hypothesis testing framework, we formulate the permutation recovery problem. The hypothesis corresponds to the true permutation (or ranking) of the data vector. Under the such hypothesis testing framework, we define decision rules for the hypothesis, and derive its optimal formula in the sense that the error probability of estimating the true permutation is minimum. This chapter is highly relevant to the following papers:

[1] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering structure of noisy data through hypothesis testing. In *2020 IEEE International Symposium on Information Theory (ISIT),* pages 1307–1312, 2020.

[2] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering data permutations from noisy observations: The linear regime. *IEEE Journal on Selected Areas in Information Theory,* 1(3):854–869, 2020.

In Chapter 3, we investigate the optimal decision rule for the data permutation recovery problem. In particular, we consider a scenario where data is generated according to an isotropic Gaussian distribution, and the perturbation consists of adding Gaussian noise that can have an arbitrary covariance matrix, i.e., noise can have memory. We show that the optimal decision regions may or may not be a linear transformation of the corresponding hypothesis regions depending on the noise covariance matrix. We focus our study on the *linear regime* where the optimal permutation decoding consists of a simple linear transformation of the noisy observation, followed by a sorting algorithm outputting the permutation along which this linear transformation is sorted. The computed linear transformation is the same for all permutations and hence, we refer to it as permutation-independent. This regime is particularly appealing as within it the optimal decoder has a complexity that is at most polynomial in n, as opposed to a brute force approach that would incur a computational complexity of $n!$.

We then characterize the optimal decision criterion for the hypothesis testing problem in the linear regime. In particular, we show that the optimal decoder declares the permutation based only on a permutation-independent linear function of the noisy observation. Our result provides both a *linear algebraic* and a *geometric* interpretations of the linear regime in terms of the noise covariance matrix. Specifically, the linear algebraic viewpoint says that the noise covariance

matrix can have at most three distinct eigenvalues. The geometric interpretation, instead says that the n-dimensional ellipsoid, characterized by a function of the noise covariance matrix, when projected onto a specific hyperplane has to be an $(n-1)$-dimensional ball. To derive these results, a core technical component consists of using linear algebraic and geometric tools, e.g., the Schur complement and Steiner symmetrization.

With the structure of the optimal decision regions in the linear regime, we discuss several practically relevant implications and special cases. For instance, we prove that when $n = 2$ the linear regime is the only regime. For the class of diagonal noise covariance matrices and $n > 2$, we show that the noise covariance matrix must have all equal diagonal elements to fall within the linear regime, i.e, if the noise is memoryless, then it must be isotropic. Finally, we characterize the error probability incurred by the decision criterion in the linear regime. In particular, we express the error probability in terms of the volume of a region that consists of the intersection of a cone with a permutation-independent linear transformation of the unit radius $2n$-dimensional ball. This chapter is highly relevant to the following papers:

[1] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering structure of noisy data through hypothesis testing. In *2020 IEEE International Symposium on Information Theory (ISIT),* pages 1307–1312, 2020.

[2] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering data permutations from noisy observations: The linear regime. *IEEE Journal on Selected Areas in Information Theory,* 1(3):854–869, 2020.

In Chapter 4, we investigate the *linear regime* discovered in Chapter 3, which determines the data permutation by applying a linear transformation to the observation followed by a sorting operation, as a decision rule. In particular, we characterize the *probability of error* of the data permutation recovery problem when a linear decoder is employed. The derived probability of error holds for any continuous data distribution, and for any Gaussian noise setting with arbitrary correlation among the entries of the noise vector. The practically relevant cases of exchangeable data distribution and isotropic noise distribution are also analyzed in detail.

We analyze the behavior of the derived probability of error in the low-noise regime. In particular, without loss of generality, we assume that the noise is isotropic, i.e., the noise has a diagonal scalar covariance matrix with σ being the noise standard deviation. In the low-noise regime (i.e., $\sigma \to 0$), the probability of error increases linearly in σ with a slope that can be a

quadratic function of n, and proportional to the L_2 norm of the input data distribution. Similar to the low-noise regime, we study the behavior of the probability of error in the high-noise regime, i.e., when $\sigma \to \infty$. We show that the error probability can be written as a weighted sum of the expected spacings between the data order statistics and that the term $1 - 1/n!$ dominates the probability of error for several distributions of interest.

We derive upper and lower bounds on the probability of correctness in terms of n for the case when $\mathbf{X}$ has i.i.d. components. Using these bounds we show that the probability of correctness decreases to zero at least exponentially fast when $n \to \infty$ (i.e., the high-dimensional regime). We also derive a universal upper bound on the probability of correctness that holds for any sub-Gaussian i.i.d. data distributions, and we provide tighter bounds for the case when $\mathbf{X}$ is i.i.d. Gaussian. This chapter is highly relevant to the following papers:

[3] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: High and low noise asymptotics. In *2021 IEEE International Symposium on Information Theory (ISIT),* pages 1100–1105, 2021.

[6] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: Asymptotics. *IEEE Transactions on Information Theory,* 70(4): 2999 –3017, 2024.

In Chapter 5, we study the *private ranking recovery problem*, where a confidential input data vector needs to be privatized (by means of a randomized mechanism) before being shared with an external party. The main objective in this chapter is to characterize the trade-off between the performance of estimating the permutation of the input data vector (measured in terms of *error probability*) and the level of *privacy* (measured in terms of ϵ-differential privacy (DP) [16] and (α, ϵ)-Rényi differential privacy (RDP) [17]) that the used mechanism guarantees.

Initially, we establish the framework for the private ranking recovery problem within the context of differential privacy. Specifically, we choose the (α, ϵ)-RDP as our metric for privacy; this preference is largely due to (α, ϵ)-RDP's ability to encompass other commonly used DP metrics, such as ϵ-DP [16] for $\alpha \to \infty$, and ϵ-KL DP [18] when $\alpha \to 1$. Additionally, as highlighted in [17, Proposition 3], it is possible to convert (α, ϵ)-RDP to (ϵ, δ)-DP. Next, we demonstrate that, assuming the input data vector has an exchangeable distribution and the randomization mechanism follows an ℓ_p-spherical distribution, the linear estimate (i.e., the permutation of noisy observation) is optimal. Subsequently, we characterize the error probability

associated with the linear decoder by employing the derivation of its Taylor series. This analysis points to the dominance of noise in the private ranking recovery problem, signifying that the probability of error remains substantial even at lower noise variances. Moreover, we present a first-order approximation of the error probability relative to the noise's standard deviation, and we validate the precision of this approximation through numerical simulations. Notably, this approximation decouples the effects of the input data and noise distributions on the error probability. Furthermore, we characterize the slope of approximated error probability, specifically for scenarios involving i.i.d. input data vector components. Lastly, we explore the trade-off between privacy (evaluated via ϵ-DP and (α, ϵ)-RDP) and utility (assessed by the error probability P_e) in the low-noise regime. In this examination, we include prevalent noise additive mechanisms such as the Laplace, Gaussian, and generalized normal mechanisms. These mechanisms have different relationships between ϵ and P_e. The analysis for both the Gaussian and Laplace methods is based on (α, ϵ)-RDP, whereas the generalized normal approach with $p \leq 1$ considers ϵ-DP. It is observed that the generalized normal mechanism, particularly when $p \leq 1$, presents an optimal balance. This chapter is relevant to the following papers:

[4] Minoh Jeong, Alex Dytso, and Martina Cardone. Ranking recovery under privacy considerations. *Transactions on Machine Learning Research,* 2022.

In Chapter 6, we investigate an *approximate* version of the ranking recovery problem. In particular, the data permutation recovery problem consists of recovering the *exact* permutation (also ranking) according to which an input data vector was sorted before being corrupted by some additive noise. We relax the exact permutation recovery problem, namely we allow for some controlled distortion in the estimation of the ranking. 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. Similar to the exact recovery problem, we first show that an optimal (in terms of error probability) decision rule for this problem is given by the *linear decoder*, which consists of simply declaring the ranking of the noisy observation. Then, 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 sub-linearly with the noise standard deviation σ. This is a notable difference with respect to the exact version of the ranking recovery problem, where we showed that the error probability grows linearly with σ. This result highlights that the 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. This chapter is relevant to the following paper:

[5] Minoh Jeong, Martina Cardone, and Alex Dytso. On the ranking recovery from noisy observations up to a distortion. In *2022 IEEE International Symposium on Information Theory (ISIT),* pages 1993–1998, 2022.

1.1 Related Work

Permutation associated estimation problems have recently gained significant importance and are studied in various fields [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 4, 41, 42, 43, 44]. The ranking (e.g., data permutation) estimation problem under a joint Gaussian distribution was investigated in [19, 20, 21, 22]. In particular, in [19] the author considered a pairwise ordering for the bivariate case; the extended version to the n-dimension was considered in [20]. The generalization of the assumption of a Gaussian distribution to an elliptically contoured distribution can be found in [21, 22]. The authors in [19, 20, 21, 22] 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. Most of recent works study the problem based on a linear regression framework, where a premultiplication by an unknown permutation matrix suitably models the problem with unknown labels. In [23], 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 [25, 24]. 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 [26]. Data estimation given randomly selected measurements – referred to as unlabeled sensing – was studied in [27, 28]. In [27], 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 [29, 30, 31]. In [32, 33], the authors studied the effect of multiple measurements on the unlabeled sensing problem and proposed an estimator using the alternating direction method of multipliers [45]. In [34], a simple one-step estimator for the unlabeled linear regression was proposed under the assumption of a

Gaussian measurement matrix and Gaussian additive noise. The estimation of a sorted vector based on noisy observations was proposed in [36], where the MMSE estimator on sorted data was characterized as a linear combination of estimators on the unsorted data. Variant versions of the unlabeled sensing problem such as sparsity, local-permutation, and principal component analysis were recently studied in [46, 47, 48, 49, 50, 51, 52].

Chapter 2

Preliminaries

This chapter provides notations and definitions used in this thesis, and the problem of permutation recovery from noisy observation is formally defined by leveraging the statistcal hypothesis testing framework. The optimal decoder, which minimizes the error probability of the hypothesis testing, is introduced under Bayesian framework that uses prior (data) and posterior distributions.

2.1 Notations

We start this chapter by listing the notations.

Table 2.1: Summary of Notations

Notation	Meaning
$\mathbf{X}$	Random vector
$\mathbf{x}$	A realization of $\mathbf{X}$
$X_{i:n}$	The i-th order statistics of $\mathbf{X} \in \mathbb{R}^n$ [53]
$[n_1 : n_2]$	Set of integers from n_1 to $n_2 \geq n_1$
I_n	Identity matrix of dimension n
$\mathbf{0}_n$	Column vector of dimension n of all zeros
$\mathbf{1}_n$	Column vector of dimension n of all ones

Continued on next page

Table 2.1 – Continued from previous page

Notation	Meaning
$0_{n\times k}$	$n \times k$ matrix of all zeros
$1_{n\times k}$	$n \times k$ matrix of all ones
$\det(A)$	Determinant of the matrix A
$\pi_{\mathbf{x}}$	The permutation sequence of $\mathbf{x}$
$\|\mathbf{x}\|_p$	ℓ_p norm of $\mathbf{x}$
$\|A\|$	For a square matrix A, the spectral norm of A
$\mathbf{x}^T$	Transpose of $\mathbf{x}$
$\mathcal{A}, \mathcal{B}, \cdots$	Calligraphic letters indicate sets
$\|\mathcal{A}\|$	Cardinality of the set $\mathcal{A}$
$\varnothing$	Empty set
$\mathbb{1}(\mathcal{A})$	The indicator function, which yields 1 if $\mathcal{A}$ is true and 0 otherwise
$\mathrm{Vol}^k(\mathcal{S})$	For a set $\mathcal{S} \subseteq \mathbb{R}^k$, the volume of $\mathcal{S}$, i.e., the k-dimensional Lebesgue measure of $\mathcal{S}$
$\mathcal{B}^n(\mathbf{c}, r)$	n-dimensional ball centered at $\mathbf{c} \in \mathbb{R}^n$ with radius r
$P_{\mathbf{X}}(\mathcal{S})$	The probability $\Pr(\mathbf{X} \in \mathcal{S})$ for some measurable set $\mathcal{S}$
$A\mathcal{B}$	The multiplication of a matrix A by a set $\mathcal{B}$, i.e., $A\mathcal{B} = \{Ax : x \in \mathcal{B}\}$
$\mathbf{x} \geq \mathbf{y}$	For two vectors $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n$, the i-th element of $\mathbf{x}$ is larger than or equal to the i-th element of $\mathbf{y}$ for all $i \in [1 : n]$.

2.2 Permutation Recovery from Noisy Observations

The problem of permutation recovery from noisy observations is a detection problem in which we seek to classify the original data vector's permutation (i.e., ordering) based on observed data vector. The problem is graphically illustrated in Fig. 2.1, where an n-dimensional random vector $\mathbf{X}$ is generated according to a certain prior distribution $P_{\mathbf{X}}$. The random vector $\mathbf{X}$ is then passed through an additive Gaussian noise channel, the output of which is denoted as $\mathbf{Y}$. In other words, 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. With the channel model, we are

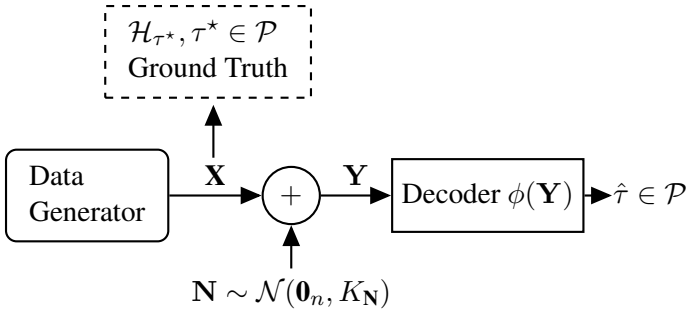


Figure 2.1: Graphical representation of the considered framework.

interested in answering the following question: *Given the observation of* $\mathbf{Y}$*, according to which permutation was the vector* $\mathbf{X}$ *sorted?*

This problem can be formulated within a hypothesis testing framework with $n!$ hypotheses $\mathcal{H}_\tau, \tau \in \mathcal{P}$, where $\mathcal{P}$ is the collection of all permutations of the elements of $[1:n]$. Specifically, $\mathcal{H}_\tau$ is the hypothesis that $\mathbf{X}$ is an n-dimensional vector sorted according to the permutation $\tau \in \mathcal{P}$; that is

$$\mathcal{H}_\tau = \{\mathbf{x} \in \mathbb{R}^n : x_{\tau_1} \leq x_{\tau_2} \leq \cdots \leq x_{\tau_n}\}, \tag{2.1}$$

with $x_{\tau_i}, i \in [1:n]$ being the τ_i-th element of $\mathbf{x}$, and $\tau_i, i \in [1:n]$ being the i-th element of τ. In terms of permutation $\pi_{\mathbf{x}}$, the hypothesis in (2.1) is equivalent to

$$\mathcal{H}_\tau = \{\mathbf{x} \in \mathbb{R}^n : \pi_{\mathbf{x}} = \tau\}. \tag{2.2}$$

Example 2.2.1. Let $n = 3$, then we have $|\mathcal{P}| = 6$ and hypotheses $\mathcal{H}_\tau, \tau \in \mathcal{P}$ 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 hypothesis region in the 3-dimensional space, as also graphically represented in Fig. 2.2.

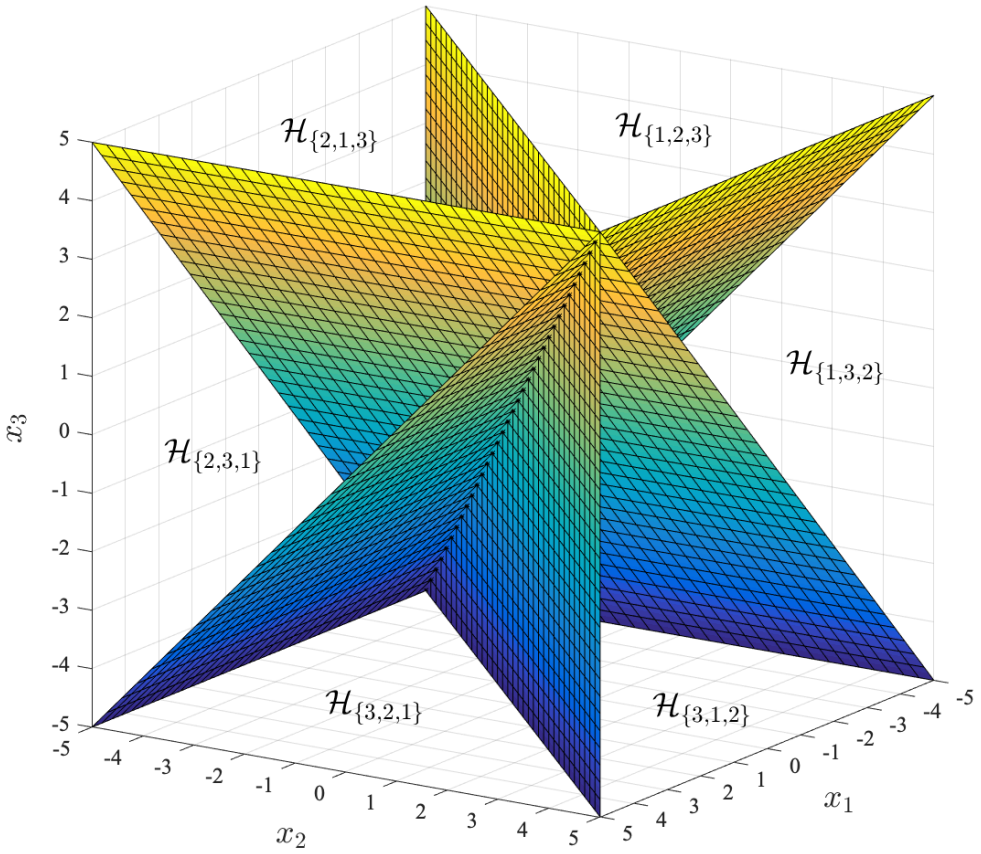


Figure 2.2: Case $n = 3$. Graphical representation of the hypothesis regions associated to each of the 6 hypotheses.

2.3 Optimal Decoder for Permutation Recovery

Given the hypothesis testing framework in (2.1), a decoder $\phi : \mathbb{R}^n \to \mathcal{P}$ in Fig. 2.1 declares the estimate of true permutation of $\mathbf{X}$. Furthermore, an *optimal* (i.e., that incurs the minimum probability of error) decoder $\phi_{\text{opt}} : \mathbb{R}^n \to \mathcal{P}$ will output $\hat{\tau} \in \mathcal{P}$ such that

$$\phi_{\text{opt}}(\mathbf{y}) = \underset{\tau \in \mathcal{P}}{\text{argmin}} \left\{ \Pr\left(\tau \neq \tau^\star \mid \mathbf{X} \in \mathcal{H}_{\tau^\star}, \mathbf{Y} = \mathbf{y}\right) \right\}, \tag{2.3}$$

where $\tau^\star$ denotes the true permutation of $\mathbf{X}$. In particular, the optimal decoder[1] will declare $\phi_{\text{opt}}(\mathbf{y}) = \hat{\tau}$ if and only if the observation vector $\mathbf{y} \in \mathcal{R}_{\hat{\tau},K_{\mathbf{N}}}$, where $\mathcal{R}_{\tau,K_{\mathbf{N}}}$'s, $\tau \in \mathcal{P}$ are the so-called *optimal* decision regions.[2] These can be derived by leveraging the maximum a posterior probability (MAP) criterion [54, Appendix 3C] as

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = \left\{ \mathbf{y} \in \mathbb{R}^n : f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) > \max_{\substack{\eta \in \mathcal{P} \\ \eta \neq \tau}} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\eta) \right\}, \tag{2.4}$$

where $f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau) P_{\mathbf{X}}(\mathcal{H}_\tau)$ with $f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau)$ denoting the conditional pdf of $\mathbf{Y}$ given that $\mathbf{X} \in \mathcal{H}_\tau$. In order to guarantee that the collection $\{\mathcal{R}_{\tau,K_{\mathbf{N}}}, \tau \in \mathcal{P}\}$ is a partition of the n-dimensional space, we assume that if $\mathbf{y} \in \{\mathcal{R}_{\tau,K_{\mathbf{N}}}, \tau \in \mathcal{S}, \mathcal{S} \subseteq \mathcal{P}, |\mathcal{S}| > 1\}$, then one of the hypotheses $\mathcal{H}_\tau, \tau \in \mathcal{S}$ is arbitrarily selected.

[1] If the argmin in (2.3) is not unique, we assume that the decoder outputs an arbitrary permutation among the candidates.

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

Chapter 3

Linear Regime of Permutation Recovery

This chapter studies the problem of permutation recovery from noisy observation. We start by considering the *exact* permutation recovery problem formulated in Chapter 2, and we investigate the optimal decoder (2.3) under Gaussian prior and additive Gaussian noise channel. The main focus of this chapter is characterizing the optimality of linear decoders in terms of parameters of Gaussian such as mean vector and covariance matrix.

3.1 Introduction

In this chapter, we investigate the permutation recovery problem defined in Section 2.2. We consider a scenario where data is generated according to an isotropic Gaussian distribution, and the perturbation consists of adding Gaussian noise that can have an arbitrary covariance matrix, i.e., noise can have memory. In other words, we have $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$ with $K_{\mathbf{N}}$ the covariance matrix of the additive noise $\mathbf{N}$, and $\mathbf{X}$ and $\mathbf{N}$ are independent. A graphical representation of the problem is given in Fig. 2.1, where we additionally assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$.

In subsequent sections, we establish that the optimal decision regions in our hypothesis testing problem exhibit a specific symmetry. Based on this insight, we illustrate that these optimal regions might or might not undergo a linear transformation from their respective hypothesis regions, contingent on the noise covariance matrix's characteristics. Our study primarily targets

the *linear regime*, characterized by an optimal permutation decoding that involves a straightforward linear transformation of the noisy data, followed by a sorting algorithm to determine the permutation. This linear transformation remains consistent across all permutations, which we refer to as permutation-independent, offering an attractive solution due to its polynomial computational complexity in n, contrary to the factorial complexity ($n!$) demanded by exhaustive methods.

We further delineate the optimal decision criterion within this linear regime, identifying the decision regions and demonstrating that the optimal decoder operates based on a permutation-independent linear function of the noisy observations. This approach yields both linear algebraic and geometric perspectives on the linear regime, influenced by the noise covariance matrix. Algebraically, it implies that the noise covariance matrix is limited to a maximum of three distinct eigenvalues. Geometrically, it suggests that an n-dimensional ellipsoid, derived from the noise covariance matrix, projects onto a hyperplane as an $(n-1)$-dimensional ball, utilizing tools such as the Schur complement and Steiner symmetrization for these derivations.

Exploring the practical implications and special cases within the linear regime, we demonstrate that for $n = 2$, the linear regime is the only regime. In the case of diagonal noise covariance matrices with $n > 2$, we prove that these matrices must possess uniformly equal diagonal elements to align with the linear regime criteria. Lastly, we characterize the error probability in the linear regime, expressing it through the volume of an intersection between a cone and a permutation-independent linear transformation of a unit radius $2n$-dimensional ball, thereby shedding light on several important practical implications and specific instances within this framework.

3.2 Optimal Decision Regions

In this section, using standard hypothesis testing tools we characterize the optimal decision criterion. We also make general statements about the structure of the decision regions. Towards this end, we make use of the result in [54, Appendix 3C], which shows that for an observation $\mathbf{y}$, the optimal decision criterion in (2.3) is given by the maximum a posterior probability (MAP)

decoder, namely

$$\mathcal{H}_{\hat{\tau}}: \ \hat{\tau} = \underset{\tau \in \mathcal{P}}{\operatorname{argmax}} \{f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau)\}, \tag{3.1a}$$

$$f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau) \Pr(\mathcal{H}_\tau), \ \tau \in \mathcal{P}, \tag{3.1b}$$

where $f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau)$ denotes the conditional probability density function (PDF) of $\mathbf{Y}$ given that $\mathbf{X} \in \mathcal{H}_\tau$. By defining the likelihood functions $L(\mathbf{y}, \mathcal{H}_\tau) = f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau), \forall \tau \in \mathcal{P}$, we have that (3.1) can be equivalently formulated as

$$\mathcal{H}_{\hat{\tau}}: \ \frac{L(\mathbf{y}, \mathcal{H}_{\hat{\tau}})}{L(\mathbf{y}, \mathcal{H}_\tau)} \geq 1, \ \forall \tau \neq \hat{\tau}, \tag{3.2}$$

where we have used the fact that $\Pr(\mathcal{H}_\tau) = \Pr(\mathcal{H}_\eta), \forall (\tau, \eta) \in \mathcal{P} \times \mathcal{P}$, 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, the likelihood function $L(\mathbf{y}, \mathcal{H}_\tau), \tau \in \mathcal{P}$ can be expressed using the convolution between two PDFs as

$$\begin{aligned} L(\mathbf{y}, \mathcal{H}_\tau) &= \int f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}|\mathcal{H}_\tau}(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= \mathbb{E}\left[f_{\mathbf{N}}(\mathbf{y} - \mathbf{X})|\mathcal{H}_\tau\right], \end{aligned} \tag{3.3}$$

where $f_{\mathbf{N}}(\cdot)$ is the PDF of $\mathbf{N}$.

With the formulation in (3.2), we can now define the *optimal* decision regions $\mathcal{R}_{\tau, K_{\mathbf{N}}}, \tau \in \mathcal{P}$ of our hypothesis testing problem[1]. The decision criterion will leverage these regions to output $\mathcal{H}_{\hat{\tau}}, \hat{\tau} \in \mathcal{P}$, namely if the observation vector $\mathbf{y} \in \mathcal{R}_{\tau, K_{\mathbf{N}}}$, the decoder would declare that the input vector $\mathbf{x} \in \mathcal{H}_\tau$. We have that the optimal decision region $\mathcal{R}_{\tau, K_{\mathbf{N}}}$ corresponding to the hypothesis region $\mathcal{H}_\tau, \tau \in \mathcal{P}$ is defined as

$$\begin{aligned} \mathcal{R}_{\tau, K_{\mathbf{N}}} &= \left\{ \mathbf{y} \in \mathbb{R}^n : f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) \geq \max_{\substack{\eta \in \mathcal{P} \\ \eta \neq \tau}} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\eta) \right\} \\ &= \left\{ \mathbf{y} \in \mathbb{R}^n : \frac{L(\mathbf{y}, \mathcal{H}_\tau)}{L(\mathbf{y}, \mathcal{H}_\eta)} \geq 1, \ \forall \eta \in \mathcal{P}, \eta \neq \tau \right\}. \end{aligned} \tag{3.4}$$

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

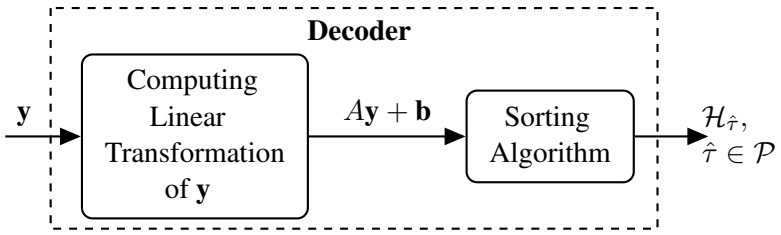


Figure 3.1: Diagram of the optimal decoder in the linear regime.

If $\mathbf{y} \in \mathbb{R}^n$ belongs to the boundary between two or more decision regions, then we arbitrarily select one of the $\mathcal{H}_\tau, \tau \in \mathcal{P}$ associated to these candidate decision regions.

3.3 Linear Regime for Optimal Decoder

We now focus to characterize *necessary and sufficient* conditions on the noise covariance matrix $K_{\mathbf{N}}$ such that each optimal decision region $\mathcal{R}_{\tau,K_{\mathbf{N}}}, \tau \in \mathcal{P}$ in (3.4) is a permutation-independent *linear* transformation of the corresponding hypothesis region $\mathcal{H}_\tau$ (i.e., $\mathcal{R}_{\tau,K_{\mathbf{N}}} = A\mathcal{H}_\tau + \mathbf{b}$ for some $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$, which are the same for all permutations). In other words, we seek to characterize the regime in which the optimal decoder consists of computing a permutation-independent linear transformation of the noisy observation $\mathbf{y}$ (i.e., $A\mathbf{y} + \mathbf{b}$), followed by a sorting algorithm outputting the permutation along which the vector $A\mathbf{y} + \mathbf{b}$ is sorted – see also Fig. 3.1. We refer to this regime as *linear*.

Characterizing the linear regime (if any) is important for several reasons. First, it is a natural step to characterizing the complete solution of the problem. Second, in the linear regime, the optimal decoder is appealing from a computational perspective. The block diagram of the optimal decoder in the linear regime is shown in Fig. 3.1. The optimal decoder first computes a permutation-independent linear transformation of $\mathbf{y}$ (first block in Fig. 3.1), which is a polynomial in n complexity task (an expression for this linear transformation is provided in Theorem 3.4.1). Next, given this linear transformation, the optimal decoder only needs to perform sorting on it (second block in Fig. 3.1), which is a task of complexity $\mathcal{O}(n \log n)$. Thus, in the linear regime, the optimal decoder has at most polynomial in n complexity. This performance should be compared to the brute force evaluation of the optimal test in (3.4), which has a practically prohibitive complexity of $n!$.

Finding a meaningful expression for the structure of $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ for all $K_{\mathbf{N}}$ seems to be a challenging task. However, some properties can be found on the structure of $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ in the general case. In particular, the following proposition, the proof of which is provided in Appendix A.1, demonstrates that the regions must have a certain symmetry. This property will also be useful for the characterization of the linear regime.

Proposition 3.3.1. *Let $(\tau,\eta) \in \mathcal{P}\times\mathcal{P}$ be the index pair that satisfies $\mathcal{H}_\tau = -\mathcal{H}_\eta$, that is $\eta_i = \tau_{n-i+1}, i \in [1:n]$, with τ_i and η_i indicating the i-th element of τ and η, respectively. Then, $\mathcal{R}_{\tau,K_{\mathbf{N}}} = -\mathcal{R}_{\eta,K_{\mathbf{N}}}$, that is for any observation $\mathbf{y} \in \mathcal{R}_{\tau,K_{\mathbf{N}}}$ it follows that $-\mathbf{y} \in \mathcal{R}_{\eta,K_{\mathbf{N}}}$.*

Remark 3.3.2. We note that the result in Proposition 3.3.1 can be generalized beyond the Gaussian assumption on $\mathbf{X} \in \mathbb{R}^n$. In particular, it holds under the condition that $\mathbf{X} \in \mathbb{R}^n$ is an exchangeable random vector.[2]

We conclude this section by providing an example of $K_{\mathbf{N}}$ that puts us outside of the linear regime. Consider $n = 3$ and the following noise covariance matrix

$$K_{\mathbf{N}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{3.5}$$

By performing brute force comparisons in (3.4), Fig. 3.2 shows the structure of the optimal decision regions for the choice of $K_{\mathbf{N}}$ in (3.5). We highlight that, for notational simplicity, in Fig. 3.2 we indicated $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ as $\mathcal{R}_\tau$. Note that the $\mathcal{H}_\tau$'s, which have a cone structure (see Fig. 2.2), cannot be a linear transformation of the $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ regions in Fig. 3.2. In Section 3.4, we will provide a formal explanation of why the covariance matrix in (3.5) does not induce a linear regime. Finally, observe that as expected, in view of Proposition 3.3.1, the optimal decision regions in Fig. 3.2 have a point of symmetry with respect to the origin.

3.4 Characteristics of the Linear Regime and Discussion

We here provide our main result on the linear regime, and we discuss several practically relevant implications of it. Our main result, which is proved in Section 3.5, provides characteristics of

[2] A sequence of random variables $U_1, U_2, \ldots, U_n$ is exchangeable if, for any permutation $(\tau_1, \tau_2, \ldots, \tau_n)$ of $[1:n]$, we have $(U_1, U_2, \ldots, U_n) \overset{d}{=} (U_{\tau_1}, U_{\tau_2}, \ldots, U_{\tau_n})$, where $\overset{d}{=}$ denotes equality in distribution.

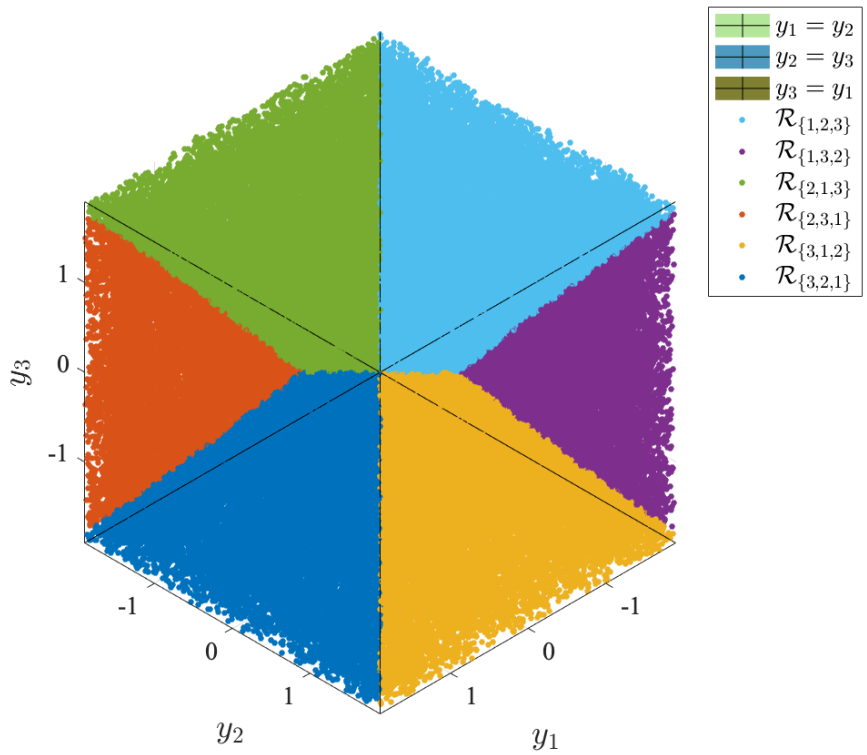


Figure 3.2: Monte Carlo simulation of the optimal decision regions $\mathcal{R}_{\tau,K_{\mathbf{N}}}, \tau \in \mathcal{P}$ where $K_{\mathbf{N}}$ is defined in (3.5).

the linear regime for optimal decoder under the isotropic Gaussian prior and the additive Gaussian channel with an arbitrary covariance matrix. In particular, the result shows geometrical properties of the optimal linear decision regions on the noise covariance matrix and proposes a closed-form expression for the decoder in terms of hypothesis region. Moreover, by showing the conditions are equivalent, it provides the necessary and sufficient conditions, under which the linear regime is attained, in terms of the noise covariance matrix $K_{\mathbf{N}}$. We then provide several implications of the conditions, such as connection to the minimum mean squared error estimator in the estimation problem (instead of permutations recovery), computational complexity.

Our main result is given by the following theorem.

Theorem 3.4.1. *The following conditions are equivalent:*

1. $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ *is a permutation-independent linear transformation of* $\mathcal{H}_\tau$*;*

2. $\mathbf{0}_n \in \bigcap_{\tau\in\mathcal{P}} \mathcal{R}_{\tau,K_{\mathbf{N}}}$;

3. *The ellipsoid* $\left(K_{\mathbf{N}}^{-1}+I_n\right)^{-\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n,1\right)$ *projected onto the hyperplane* $\mathcal{W}=\{\mathbf{x}\in\mathbb{R}^n : \mathbf{1}_n^T\mathbf{x}=0\}$ *is an* $(n-1)$*-dimensional ball of radius* γ *for some constant* $\gamma\in(0,1)$*;*

4. *Let* $\mathcal{Q}=\left\{Q\in\mathcal{SO}(n) : \mathbf{q}_n=\frac{1}{\sqrt{n}}\mathbf{1}_n\right\}$*, where* $\mathcal{SO}(n)$ *is the set of* $n\times n$ *real-valued orthonormal matrices, and* $\mathbf{q}_n$ *is the* n*-th column of* Q*. Then,*

$$\left(K_{\mathbf{N}}^{-1}+I_n\right)^{-1}=Q\begin{bmatrix}\gamma I_{n-2} & 0_{n-2\times 2}\\ 0_{2\times n-2} & S\end{bmatrix}Q^T, \tag{3.6}$$

where $Q\in\mathcal{Q}$*,* $S=\left[\begin{smallmatrix}\gamma & v\\ v & a\end{smallmatrix}\right]$ *and* $\gamma\in(0,1)$*,* $a\in(0,1)$*,* $v\in\mathbb{R}$ *such that* $v^2<\min\{a\gamma,(1-a)(1-\gamma)\}$*; and*

5. $\mathcal{R}_{\tau,K_{\mathbf{N}}}=\left(K_{\mathbf{N}}+I_n\right)\mathcal{H}_\tau$*, for all* $\tau\in\mathcal{P}$*.*

Remark 3.4.2. Recall that for $\mathbf{X}\sim\mathcal{N}(\mathbf{0}_n,I_n)$, we have that

$$\mathbf{X}|\mathbf{Y}=\mathbf{y}\sim\mathcal{N}\left(\mathbb{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right],\text{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})\right),$$

where $\mathbb{E}\left[\mathbf{X}|\mathbf{Y}=\mathbf{y}\right]=(I_n+K_{\mathbf{N}})^{-1}\mathbf{y}$ [55] and $\text{Var}(\mathbf{X}|\mathbf{Y}=\mathbf{y})=(I_n+K_{\mathbf{N}}^{-1})^{-1},\ \forall\mathbf{y}\in\mathbb{R}^n$. It therefore follows that condition 4) in Theorem 3.4.1 imposes a constraint for the conditional covariance of $\mathbf{X}$ given $\mathbf{Y}$. Moreover, recall that the conditional expectation is the optimal mean squared error estimator [55]. Therefore, the permutation-independent linear transformation in condition 5) in Theorem 3.4.1 is, in fact, the optimal linear estimator – see also first block in Fig. 3.1.

Remark 3.4.3. One interesting property of condition 4) in Theorem 3.4.1 is the following. Let $\mathbf{G}=\mathbf{X}|\mathbf{Y}$ be the Gaussian random vector that has properties as indicated in Remark 3.4.2. Then, it can be shown that $\left\{\frac{1}{i}\sum_{k=1}^{i}G_k-G_{i+1}\right\}, i\in[1:n-1]$ are independent. In particular, this follows by studying $Q^T\mathbf{G}$ that has covariance given by

$$Q^TK_{\mathbf{G}}Q=\begin{bmatrix}\gamma I_{n-2} & 0_{(n-2)\times 2}\\ 0_{2\times(n-2)} & S\end{bmatrix}, \tag{3.7}$$

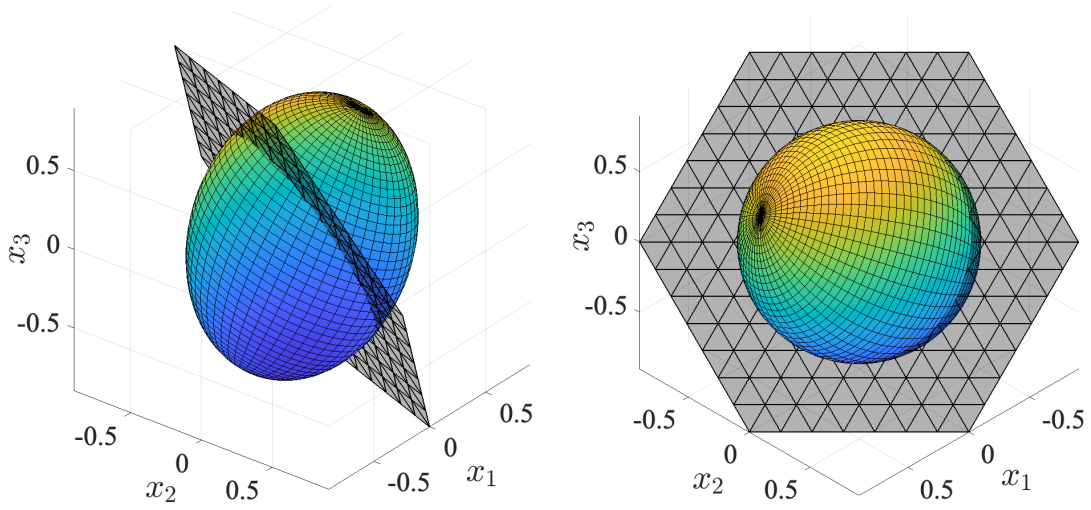


Figure 3.3: Graphical representation of the ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$, where $K_{\mathbf{N}}$ satisfies (3.6) with parameters defined in (3.8).

where $Q \in \mathcal{Q}$ is chosen such that its element $Q_{i,j}$ in the i-th row and j-th column is

$$Q_{i,j} = \begin{cases} (j^2 + j)^{-\frac{1}{2}}, & j \neq n,\ i \leq j, \\ -(1 + j^{-1})^{-\frac{1}{2}}, & j \neq n,\ i = j + 1, \\ n^{-\frac{1}{2}}, & j = n, \\ 0, & \text{otherwise.} \end{cases}$$

Remark 3.4.4. As discussed in Section 3.2, the computational complexity of the optimal decoder in the linear regime is at most polynomial in n. It is also interesting to comment on the computational complexity of verifying whether a given $K_{\mathbf{N}}$ induces a linear regime. Observe that the linearity condition in (3.6) requires to perform matrix inversion, multiplication, and eigendecomposition. All these are polynomial in n complexity tasks. Therefore, verifying if the given $K_{\mathbf{N}}$ satisfies (3.6) is a polynomial in n complexity task.

An example of $K_{\mathbf{N}}$ that induces the linear regime can be obtained by considering $n = 3$ and

$$(\gamma, a, v) = (0.5, 0.5, 0.2) \tag{3.8}$$

in (3.6). By taking the eigendecomposition of this $K_{\mathbf{N}}$, it can be verified that it has three distinct eigenvalues given by $\lambda_1 = 1$, $\lambda_2 = 3/7$ and $\lambda_3 = 7/3$. The corresponding ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ has three distinct radii, and it is shown in Fig. 3.3 (left). The projection of this ellipsoid onto $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{1}_3^T \mathbf{x} = 0\}$ is equal to a 2-dimensional ball of radius $\gamma = 1/2$ as also illustrated in Fig. 3.3 (right).

Fig. 3.4 shows that the corresponding optimal decision regions $\mathcal{R}_{\tau,K_{\mathbf{N}}}, \tau \in \mathcal{P}$, are indeed obtained as a permutation-independent linear transformation of the corresponding hypothesis regions in Fig. 2.2, namely as $\mathcal{R}_{\tau,K_{\mathbf{N}}} = \left(K_{\mathbf{N}} + I_3\right) \mathcal{H}_\tau$. We highlight that, for notational simplicity, in Fig. 3.4 we indicated $\mathcal{R}_{\tau,K_{\mathbf{N}}}$ as $\mathcal{R}_\tau$.

3.4.1 Sufficient and Necessary Conditions on the Spectrum and on the Eigenvectors of $K_{\mathbf{N}}$

We here provide sufficient and necessary conditions on the spectrum of $K_{\mathbf{N}}$, i.e., on the set of its eigenvalues, as well as on its eigenvectors that need to be satisfied for (3.6) to hold. In particular, we have the next proposition, the proof of which can be found in Appendix A.2.

Proposition 3.4.5. *A $K_{\mathbf{N}}$ satisfies the condition in* (3.6) *if and only if it has eigenvalues $\lambda_i,\ i \in [1:n]$ and eigenvectors $\boldsymbol{\nu}_i,\ i \in [1:n]$ that are in either one of the two forms below:*

- *Case 1: All the n eigenvalues are the same; we have*

$$\lambda_i = \frac{\gamma}{1-\gamma}, \qquad \boldsymbol{\nu}_i = \mathbf{t}_i, \tag{3.9}$$

  where $\{\mathbf{t}_i,\ i \in [1:n]\}$ is any set of orthogonal vectors in $\mathbb{R}^n$, and $\gamma \in (0,1)$;

- *Case 2: At least two eigenvalues are different; we have*

$$\lambda_i = \begin{cases} \frac{\gamma}{1-\gamma}, & i \in [1:n-2], \\ \frac{a+\gamma+\sqrt{(a-\gamma)^2+4v^2}}{2-a-\gamma-\sqrt{(a-\gamma)^2+4v^2}}, & i = n-1, \\ \frac{a+\gamma-\sqrt{(a-\gamma)^2+4v^2}}{2-a-\gamma+\sqrt{(a-\gamma)^2+4v^2}}, & i = n, \end{cases} \tag{3.10a}$$

$$\boldsymbol{\nu}_i = \begin{cases} \mathbf{q}_i, & i \in [1:n-2], \\ \left(v + a - \frac{\lambda_i}{1+\lambda_i}\right) \mathbf{q}_{n-1} & \\ \quad + \left(v + \gamma - \frac{\lambda_i}{1+\lambda_i}\right) \mathbf{q}_n, & i \in [n-1:n], \end{cases} \tag{3.10b}$$

where $\gamma \in (0,1)$*,* $a \in (0,1)$*,* $v \in \mathbb{R}$ *satisfying* $v^2 < \min\{a\gamma, (1-a)(1-\gamma)\}$*, and* $\mathbf{q}_i, i \in [1:n]$ *is the* i*-th column of* $Q \in \mathcal{Q}$*.*

Remark 3.4.6. Proposition 3.4.5 provides sufficient and necessary conditions for $K_{\mathbf{N}}$ to satisfy Theorem 3.4.1 in terms of its eigenvalues and eigenvectors. Specifically, Proposition 3.4.5 shows that a $K_{\mathbf{N}}$ that satisfies (3.6) has at most three distinct eigenvalues.

3.4.2 Case of $n = 2$ is Special

It is interesting to note that in the case of $n = 2$ the condition in (3.6) is not restrictive, i.e., all covariance matrices satisfy (3.6). To put it in other words, for $n = 2$ the linear regime is the only regime, and Theorem 3.4.1 gives a complete characterization of the permutation recovery problem.

One intuitive explanation why this follows is given by condition 3) in Theorem 3.4.1 which requires that the projection of an n-dimensional ellipsoid onto the hyperplane $\mathcal{W}$ is an $(n-1)$-dimensional ball. When $n = 2$, this corresponds to projecting an ellipse onto a line. The result of this operation is a segment, which is indeed a 1-dimensional ball. Therefore, for the case of $n = 2$ any $K_{\mathbf{N}}$ satisfies (3.6). We next prove this formally using condition 4) in Theorem 3.4.1.

Proposition 3.4.7. *Let* $n = 2$*. Then, every positive definite covariance matrix* $K_{\mathbf{N}}$ *satisfies* (3.6)*.*

Proof. For $n = 2$ and any positive definite symmetric $K_{\mathbf{N}}$, the left-hand side of (3.6) can be represented by (w, q, z) as

$$(K_{\mathbf{N}}^{-1} + I_n)^{-1} = \begin{bmatrix} w & q \\ q & z \end{bmatrix}, \tag{3.11}$$

where $w > 0$, $z > 0$, and $wz > q^2$. Note also that the eigenvalues of the left-hand side of (3.11) are smaller than one, and hence the triple (w, q, z) has also to satisfy this constraint. Hence, we would need to find a triple (a, γ, v) such that

$$\begin{bmatrix} w & q \\ q & z \end{bmatrix} = Q \begin{bmatrix} \gamma & v \\ v & a \end{bmatrix} Q^T, \tag{3.12}$$

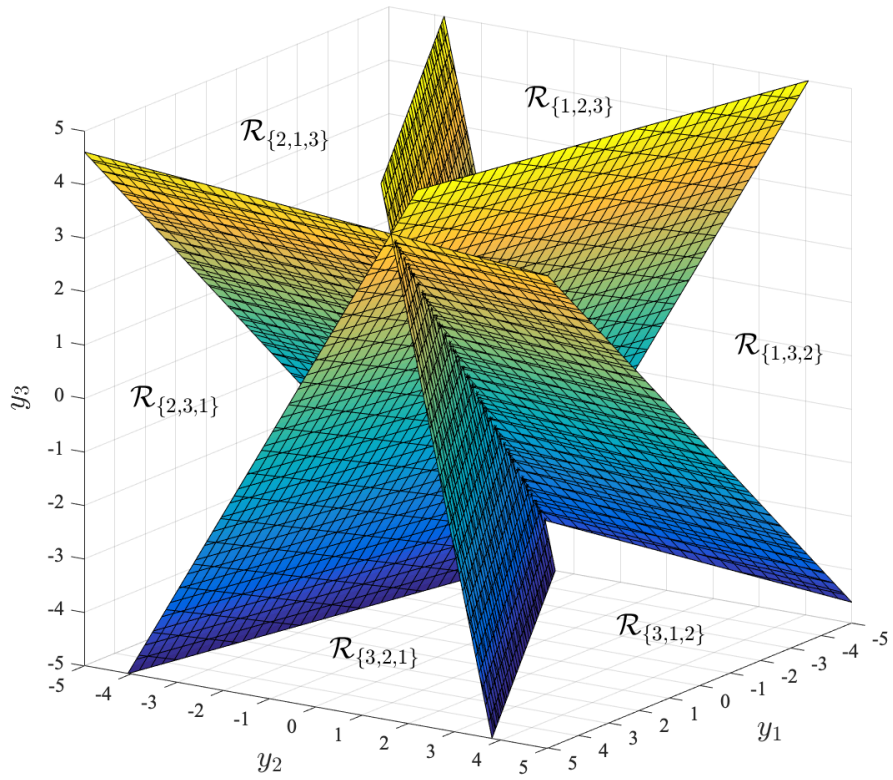


Figure 3.4: Optimal decision regions of the $K_{\mathbf{N}}$ that satisfies (3.6) with parameters defined in (3.8).

where the orthonormal matrix $Q \in \mathcal{Q}$ can be chosen as

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{3.13}$$

It is not difficult to see that the triple (a, γ, v) such that

$$a = \frac{w + z + 2q}{2}, \quad \gamma = \frac{w + z - 2q}{2}, \quad v = \frac{z - w}{2},$$

satisfies all the constraints in condition 4) of Theorem 3.4.1. This concludes the proof of Proposition 3.4.7. □

3.4.3 For $n > 2$ Memoryless Noise Can Only be Isotropic

We here focus on the case $n > 2$, and we prove that if the noise is memoryless, i.e., $K_{\mathbf{N}}$ is a diagonal matrix, then all its diagonal elements have to be equal to ensure that (3.6) is satisfied, i.e., the noise has to be isotropic. We note that this result justifies the fact that the $K_{\mathbf{N}}$ in (3.5) puts us outside of the linear regime (see Fig. 3.2). We also highlight that such a restriction does not apply for the case $n = 2$ since, as we have shown in Proposition 3.4.7, for this case any $K_{\mathbf{N}}$ satisfies (3.6).

Proposition 3.4.8. *Consider $n > 2$ and let $K_{\mathbf{N}}$ be a diagonal positive definite matrix. Then, $K_{\mathbf{N}}$ satisfies* (3.6) *if and only if*

$$K_{\mathbf{N}} = \frac{\gamma}{1-\gamma} I_n, \tag{3.14}$$

for some $\gamma \in (0, 1)$.

Proof. Let $K_{\mathbf{N}} \in \mathbb{R}^{n\times n}, n > 2$ be a diagonal matrix with $\sigma_i^2,\ i \in [1 : n]$ in its diagonal entries. We start by noting that if $K_{\mathbf{N}}$ is isotropic (i.e., $K_{\mathbf{N}} = cI_n$ for any constant $c > 0$), then it has eigenvalues λ_i and eigenvectors $\boldsymbol{\nu}_i$ as in (3.9). Thus, if $K_{\mathbf{N}}$ is isotropic, then it satisfies the condition in (3.6).

We now show that any diagonal positive definite $K_{\mathbf{N}}$ has to be of the form as in (3.14) to satisfy (3.6). To this end, assume that $K_{\mathbf{N}}$ is non-isotropic. For $i \in [1 : n]$, since $K_{\mathbf{N}}$ is a diagonal matrix, it has eigenvalues λ_i and eigenvectors $\boldsymbol{\nu}_i$ given by

$$\lambda_i = \sigma_i^2, \qquad \boldsymbol{\nu}_i = \mathbf{e}_i, \tag{3.15}$$

where $\mathbf{e}_i \in \mathbb{R}^n$ denotes an n-dimensional vector of all-zeros except a non-zero element in the i-th position. However, from Proposition 3.4.5, we know that there exists $i \in [1 : n]$ for which $\boldsymbol{\nu}_i = \frac{\gamma-a}{\sqrt{n}}\mathbf{1}_n$ (since $v = 0$), and hence a $K_{\mathbf{N}}$ that is diagonal, but non-isotropic does not satisfy the condition in (3.6). This concludes the proof of Proposition 3.4.8. □

3.4.4 On the Probability of Error

For the probability of error, we make a few comments about it. Specifically, the structure of the optimal decision regions in Theorem 3.4.1 can now be utilized to provide the following geometric characterization of the error probability, the proof of which can be found in Appendix A.3.

Proposition 3.4.9. *Let $K_{\mathbf{N}}$ satisfy the conditions in Theorem 3.4.1. Then, the error probability is given by*

$$P_e = 1 - n! \frac{\mathrm{Vol}^{2n}\left(\mathcal{C}_{\mathcal{H}_\tau} \cap A\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}{\det\left(K_{\mathbf{N}}^{\frac{1}{2}}\right) \mathrm{Vol}^{2n}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}, \tag{3.16a}$$

where

$$A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & K_{\mathbf{N}}^{\frac{1}{2}} \end{bmatrix}, \quad \mathcal{C}_{\mathcal{H}_\tau} = \mathcal{H}_\tau \times (K_{\mathbf{N}} + I_n)\mathcal{H}_\tau, \tag{3.16b}$$

and where $\tau \in \mathcal{P}$ can be chosen arbitrarily.

The result in Proposition 3.4.9 can now be used to derive various upper and lower bounds on the probability of error, and hence find *impossibility* results, i.e., properties on the noise covariance matrix $K_{\mathbf{N}}$ for which reasonable recovery is not possible.

3.4.5 Discussion on Possible Extensions

We discuss a few possible future directions and extensions. Perhaps one of the most natural next directions is to look beyond the linear regime. For example, it would be interesting to understand whether the optimal decoder always has a reasonable closed-form characterization. In particular, Proposition 3.3.1 and the results in Fig. 3.2 suggest that the optimal decision regions have a symmetrical polyhedral structure, and it would be interesting to see if this structure can be characterized. The possibility that such a characterization exists stems from the following characterization of the optimal decoder: given an observation $\mathbf{y}$

$$\begin{aligned} \tau^\star &= \arg\max_{\tau\in\mathcal{P}} \Pr[\mathbf{X} \in \mathcal{H}_\tau | \mathbf{Y} = \mathbf{y}] \\ &= \arg\max_{\tau\in\mathcal{P}} \Pr[(I_n + K_{\mathbf{N}})^{-1}\mathbf{y} + (I_n + K_{\mathbf{N}}^{-1})^{-\frac{1}{2}}\mathbf{Z} \in \mathcal{H}_\tau], \end{aligned} \tag{3.17}$$

where $\mathbf{Z}$ is a standard Gaussian random vector. The proof of the second equality in (3.17) follows from the fact that $\mathbf{X}$ given $\mathbf{Y}$ is Gaussian; see Remark 3.4.2 for more details.

It would also be interesting to study the probability of error for the linear decoder proposed in this work and compare it with the probability of error of the optimal decoder in the regimes not covered by Theorem 3.4.1. Recall that the optimal decoder in the linear regime consists of the optimal linear estimator combined with a sorting operation (see Remark 3.4.2 and Fig. 3.1). This decoder is very attractive as it is relatively easy to implement in practice. In particular, it is reasonable to suspect that there exists a large set of noise covariance matrices for which such a decoder will perform relatively well.

Another interesting direction is to consider whether the results of this paper can be generalized beyond the assumption that $\mathbf{X} \in \mathbb{R}^n$ is Gaussian. One attractive direction to consider is the case when $\mathbf{X} \in \mathbb{R}^n$ is exchangeable. The assumption of exchangeability still allows to use the symmetry argument, and in particular, Proposition 3.3.1 holds under this assumption (see Remark 3.3.2). Furthermore, let $\mathbf{X}_\mathbf{y} \in \mathbb{R}^n$ be the random variable distributed according to $f_{\mathbf{X}|\mathbf{Y}}(\cdot|\mathbf{y})$; then, from Proposition 3.3.1 it follows that the linear regime is optimal if and only if there exists a constant $c \in (0, 1)$ such that

$$\Pr[\mathbf{X}_0 \in \mathcal{H}_\tau] = c, \forall \tau \in \mathcal{P}. \tag{3.18}$$

Thus, an interesting future direction would consist of identifying the family of the noise covariance matrices for which (3.18) holds when $\mathbf{X} \in \mathbb{R}^n$ is exchangeable, but not necessarily Gaussian. Some initial steps toward this topic can be found in Chapter 4.

3.5 Proof of Theorem 3.4.1

In this section, we prove the results in Theorem 3.4.1. In particular, the proof follows the next sequence of implications

$$1) \Rightarrow 2) \Leftrightarrow 3) \Leftrightarrow 4) \Rightarrow 5) \Rightarrow 1),$$

which are next analyzed in different subsections. Note that the implication $5) \Rightarrow 1)$ follows immediately.

3.5.1 Proof of the Implication $1) \Rightarrow 2)$

We prove that $1) \Rightarrow 2)$, i.e., the fact that $\mathcal{R}_{\tau,K_\mathbf{N}}$ is a permutation-independent linear transformation of $\mathcal{H}_\tau$ implies that $\mathbf{0}_n \in \bigcap_{\tau\in\mathcal{P}} \mathcal{R}_{\tau,K_\mathbf{N}}$. To this end, we prove the next lemma by leveraging

the symmetry condition proved in Proposition 3.3.1.

Lemma 3.5.1. *Suppose that*

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = A\mathcal{H}_\tau + \mathbf{b},\ \forall \tau \in \mathcal{P}, \tag{3.19}$$

where A is an $n \times n$ matrix, and $\mathbf{b}$ is an n-dimensional column vector. Then, $\mathbf{0}_n \in \bigcap_{\tau\in\mathcal{P}} \mathcal{R}_{\tau,K_{\mathbf{N}}}$. Moreover, $\mathbf{b}$ must be of the form $\mathbf{b} = tA\mathbf{1}_n$ for some $t \in \mathbb{R}$.

Proof. Let $\mathcal{L}_\mathcal{H} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in \bigcap_{\tau\in\mathcal{P}} \mathcal{H}_\tau\right\}$ be the set of points that belong to the intersection of $\mathcal{H}_\tau,\ \forall \tau \in \mathcal{P}$. Note that this set of points forms a line in $\mathbb{R}^n$, which is given by

$$\mathcal{L}_\mathcal{H} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \kappa\mathbf{1}_n,\ \kappa \in \mathbb{R}\right\}. \tag{3.20}$$

Similarly, let $\mathcal{L}_\mathcal{R} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in \bigcap_{\tau\in\mathcal{P}} \mathcal{R}_{\tau,K_{\mathbf{N}}}\right\}$ be the set of points that belong to the intersection of $\mathcal{R}_{\tau,K_{\mathbf{N}}},\ \forall \tau \in \mathcal{P}$. Note that this set is non-empty. From the assumption in Lemma 3.5.1, we have that $\mathcal{L}_\mathcal{R} = A\mathcal{L}_\mathcal{H} + \mathbf{b}$. Thus, $\mathcal{L}_\mathcal{R}$ is also a line in $\mathbb{R}^n$ defined as

$$\mathcal{L}_\mathcal{R} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \kappa A\mathbf{1}_n + \mathbf{b},\ \kappa \in \mathbb{R}\right\}. \tag{3.21}$$

Now let $\mathbf{0}_n \neq \tilde{\mathbf{y}} \in \mathcal{L}_\mathcal{R}$. Then, by Proposition 3.3.1 if $\tilde{\mathbf{y}} \in \mathcal{L}_\mathcal{R}$, we have that $-\tilde{\mathbf{y}} \in \mathcal{L}_\mathcal{R}$. Since $\mathcal{L}_\mathcal{R}$ is a line that contains both $-\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}$, it must contain also $\mathbf{0}_n$. Finally, observe that the only $\mathbf{b}$ that is allowed (i.e., that ensures that the line contains both $-\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}$) is of the form $\mathbf{b} = tA\mathbf{1}_n$ for some $t \in \mathbb{R}$. This concludes the proof of Lemma 3.5.1. □

Note that the fact that the shift vector $\mathbf{b}$ in Lemma 3.5.1 is of the form $\mathbf{b} = tA\mathbf{1}_n$, for some $t \in \mathbb{R}$, implies that

$$\mathcal{L}_\mathcal{R} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = (\kappa + t)A\mathbf{1}_n,\ \kappa, t \in \mathbb{R}\right\} = A\mathcal{L}_\mathcal{H}, \tag{3.22}$$

and

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = A\mathcal{H}_\tau + \mathbf{b} = A(\mathcal{H}_\tau + t\mathbf{1}_n) = A\mathcal{H}_\tau. \tag{3.23}$$

In other words, such a choice of $\mathbf{b}$ does not affect the shape of the decision regions.

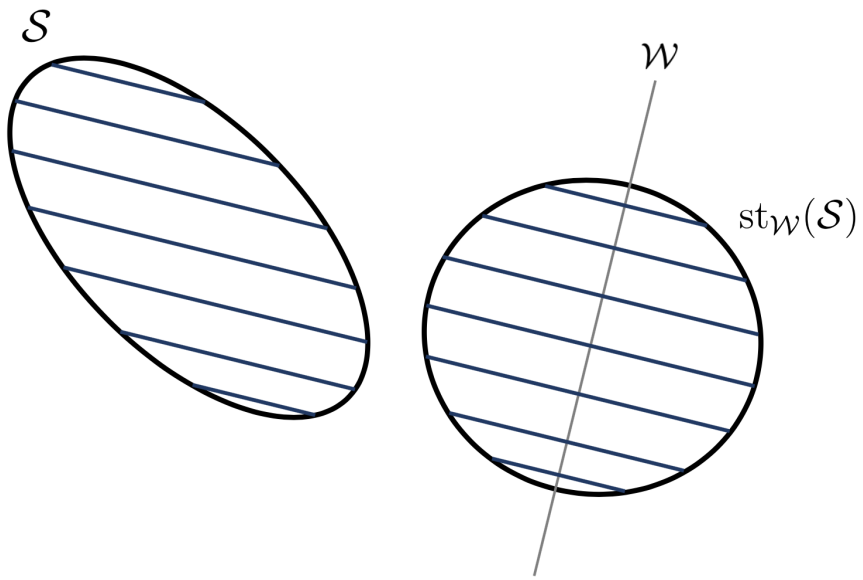


Figure 3.5: Steiner symmetrization.

3.5.2 Proof of the Implication $2) \Leftrightarrow 3)$

We here prove that $2) \Leftrightarrow 3)$, i.e., the fact that $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ projected onto $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T \mathbf{x} = 0\}$ is an $(n-1)$-dimensional ball of radius γ for some $\gamma \in (0,1)$ implies that $\mathbf{0}_n \in \bigcap_{\tau \in \mathcal{P}} \mathcal{R}_{\tau, K_{\mathbf{N}}}$, and vice versa.

In particular, the proofs $2) \Leftarrow 3)$ and $2) \Rightarrow 3)$ will leverage a symmetrization method known as Steiner symmetrization [56], which we next formally define.

Definition 3.5.2. Let $\mathcal{S}$ be a bounded set in $\mathbb{R}^n$, and $\mathcal{W}$ be an $(n-1)$-dimensional vector subspace of $\mathbb{R}^n$. The Steiner symmetrization of $\mathcal{S}$ with respect to $\mathcal{W}$ is the operation that associates the set $\mathrm{st}_{\mathcal{W}}(\mathcal{S})$ in $\mathbb{R}^n$ to the set $\mathcal{S}$ such that, for each straight line ℓ perpendicular to $\mathcal{W}$, we have that $\ell \cap st_{\mathcal{W}}(\mathcal{S})$ is either a closed line segment with the center in $\mathcal{W}$ or is empty. Moreover, the two following conditions need to be satisfied

$$\mathsf{length}\left(\ell \cap \mathcal{S}\right) = \mathsf{length}\left(\ell \cap \mathrm{st}_{\mathcal{W}}(\mathcal{S})\right), \tag{3.24a}$$

and

$$\ell \cap \mathrm{st}_{\mathcal{W}}(\mathcal{S}) = \varnothing \quad \text{if and only if} \quad \ell \cap \mathcal{S} = \varnothing. \tag{3.24b}$$

Fig. 3.5 illustrates the application of Steiner symmetrization on the set $\mathcal{S}$ with respect to the

line $\mathcal{W}$. We now provide some properties of Steiner symmetrization that will be useful in the upcoming proofs.

Proposition 3.5.3. *The Steiner symmetrization* $\mathrm{st}_{\mathcal{W}}(\mathcal{S})$ *of the set* $\mathcal{S}$ *with respect to* $\mathcal{W}$ *satisfies the following properties:*

- *Steiner symmetrization preserves convexity. Moreover, it transforms ellipsoids into ellipsoids [57].*
- *Steiner symmetrization preserves the volume, i.e.,* $\mathrm{Vol}^n(\mathcal{S}) = \mathrm{Vol}^n(\mathrm{st}_{\mathcal{W}}(\mathcal{S}))$ *[56].*
- *Steiner symmetrization preserves the orthogonal projection onto* $\mathcal{W}$*, i.e.,* $\mathrm{Proj}_{\mathcal{W}}(\mathcal{S}) = \mathrm{Proj}_{\mathcal{W}}(\mathrm{st}_{\mathcal{W}}(\mathcal{S}))$*, where* $\mathrm{Proj}_{\mathcal{W}}(\mathcal{A})$ *denotes the orthogonal projection of the set* $\mathcal{A}$ *onto* $\mathcal{W}$ *[58].*

Another result that we will leverage to prove $2) \Leftrightarrow 3)$ is provided by the following lemma, the proof of which can be found in Appendix A.4.

Lemma 3.5.4. *Let* $\mathbf{U} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{U}})$*, where* $K_{\mathbf{U}}$ *is positive definite. Then,*

$$\Pr(\mathbf{U} \in \mathcal{H}_\tau) = \frac{\left|\det\left(K_{\mathbf{U}}^{-\frac{1}{2}}\right)\right| \mathrm{Vol}^n\left(\mathcal{H}_\tau \cap K_{\mathbf{U}}^{\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)\right)}{\mathrm{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}. \tag{3.25}$$

We are now ready to prove $2) \Leftrightarrow 3)$, the proof of which consists of two parts. The first part is provided in the next lemma, which leverages the observation in Remark 3.4.2 and is proved in Appendix A.5.

Lemma 3.5.5. $\mathbf{0}_n \in \bigcap_{\tau \in \mathcal{P}} \mathcal{R}_{\tau, K_{\mathbf{N}}}$ *if and only if there exists a constant* $\eta > 0$ *such that*

$$\mathrm{Vol}^n\left(\mathcal{H}_\tau \cap \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right) = \eta, \ \forall \tau \in \mathcal{P}. \tag{3.26}$$

The second part of the proof $2) \Leftrightarrow 3)$ is given by the next lemma, which characterizes the solution of (3.26) in terms of $K_{\mathbf{N}}$ and relies on the Steiner symmetrization technique.

Lemma 3.5.6. *A* $K_{\mathbf{N}}$ *is a solution for* (3.26) *if and only if there exists a constant* $\gamma \in (0, 1)$ *such that the ellipsoid* $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ *projected onto the hyperplane* $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T \mathbf{x} = 0\}$ *is an* $(n-1)$*-dimensional ball of radius* γ*.*

Proof. Let $\mathcal{L}_{\mathcal{H}} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in \bigcap_{\tau\in\mathcal{P}} \mathcal{H}_\tau\right\}$ be the set of points that belong to the intersection of $\mathcal{H}_\tau,\ \forall\tau \in \mathcal{P}$. From (3.20), we have that

$$\mathcal{L}_{\mathcal{H}} = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \kappa\mathbf{1}_n,\ \kappa \in \mathbb{R}\right\}, \tag{3.27}$$

which is a line in $\mathbb{R}^n$. From Lemma 3.5.5, we have that $\mathbf{y} = \mathbf{0}_n$ is a boundary point for all the optimal decision regions, i.e., $\mathbf{0}_n \in \bigcap_{\tau\in\mathcal{P}} \mathcal{R}_{\tau,K_{\mathbf{N}}}$, if and only if

$$\mathrm{Vol}^n\left(\mathcal{H}_\tau \cap \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right) = \eta,\ \forall\tau \in \mathcal{P}, \tag{3.28}$$

for some $\eta > 0$. In particular, with reference to (3.28), $\mathcal{H}_\tau$ is an n-dimensional cone, and $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)$ is an n-dimensional ellipsoid centered at $\mathbf{0}_n$. We also highlight that $\mathcal{H}_\tau,\ \forall\tau$ are all open sets along the direction $\mathcal{L}_{\mathcal{H}}$, i.e., for any $\tau \in \mathcal{P}$ and $\kappa \in \mathbb{R}$, if $\tilde{\mathbf{x}} \in \mathcal{H}_\tau$, then $\tilde{\mathbf{x}} + \kappa\mathbf{1}_n \in \mathcal{H}_\tau$.

For ease of geometrical representation, we now apply Steiner symmetrization (see Definition 3.5.2) on the ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)$. In particular, with reference to Definition 3.5.2, we consider Steiner symmetrization with respect to the hyperplane

$$\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T\mathbf{x} = 0\}, \tag{3.29}$$

which is perpendicular to the line $\mathcal{L}_{\mathcal{H}}$ in (3.27). Note that $\mathcal{W}$ is an $(n-1)$-dimensional vector subspace of $\mathbb{R}^n$. By applying Steiner symmetrization on the ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)$ with respect to $\mathcal{W}$ in (3.29), we obtain a new ellipsoid $\mathcal{E}^n$ (see Proposition 3.5.3) given by

$$\mathcal{E}^n = \mathrm{st}_{\mathcal{W}}\left(\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)\right), \tag{3.30}$$

which has the same volume of the original ellipsoid (see Proposition 3.5.3), namely

$$\mathrm{Vol}^n\left(\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)\right) = \mathrm{Vol}^n\left(\mathcal{E}^n\right).$$

It is also worth noting that $\mathcal{E}^n$ is centered at $\mathbf{0}_n$, it has $\mathcal{L}_{\mathcal{H}}$ in (3.27) as an axis, and it is symmetric with respect to $\mathcal{W}$. These properties, together with the fact that $\mathcal{H}_\tau$'s with $\tau \in \mathcal{P}$ are all open

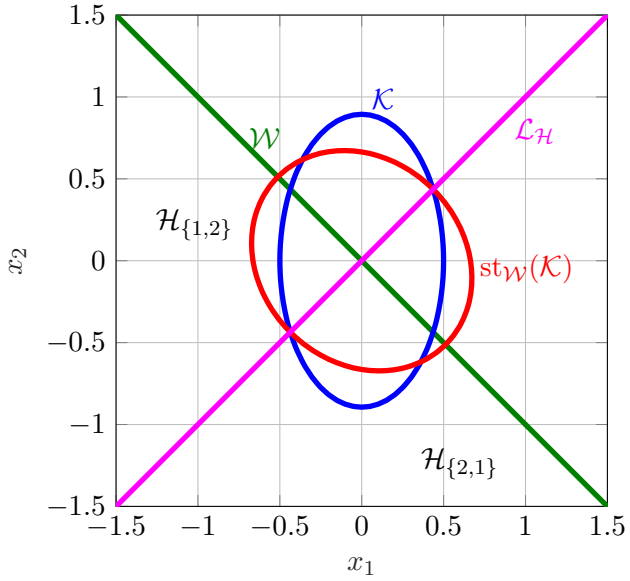


Figure 3.6: Steiner symmetrization of the ellipsoid $\mathcal{K} = \left(K_{\mathbf{N}}^{-1} + I_2\right)^{-\frac{1}{2}} \mathcal{B}^2(\mathbf{0}_2, 1)$ with respect to $\mathcal{W}$ in (3.29) where $K_{\mathbf{N}} = \left[\begin{smallmatrix} \frac{1}{3} & 0 \\ 0 & 4 \end{smallmatrix}\right]$.

sets along the direction $\mathcal{L}_{\mathcal{H}}$, imply that

$$\mathrm{Vol}^n\left(\mathcal{H}_\tau \cap \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)\right) = \mathrm{Vol}^n(\mathcal{H}_\tau \cap \mathcal{E}^n). \tag{3.31}$$

A graphical representation of the procedure explained above is given in Fig. 3.6 for the 2-dimensional case. From the analysis above, it hence follows that the problem of finding the family of $K_{\mathbf{N}}$'s that satisfies (3.28) is equivalent to finding the family of $K_{\mathbf{N}}$'s such that there exists a constant $\eta > 0$ for which

$$\mathrm{Vol}^n\left(\mathcal{H}_\tau \cap \mathcal{E}^n\right) = \eta, \ \forall \tau \in \mathcal{P}. \tag{3.32}$$

We now leverage the following lemma, the proof of which can be found in Appendix A.6, which provides sufficient and necessary conditions for (3.32) to hold.

Lemma 3.5.7. *Let $\mathcal{E}^n$ be an n-dimensional ellipsoid centered at the origin and having one axis*

of the type $\boldsymbol{\nu} = \frac{1}{\sqrt{n}}\mathbf{1}_n$*. Then, there exists* $\eta > 0$*, such that*

$$\mathrm{Vol}^n\left(\mathcal{H}_\tau \cap \mathcal{E}^n\right) = \eta,\ \forall \tau \in \mathcal{P}, \tag{3.33}$$

if and only if $\mathcal{E}^n$ *has equal radii for all axes except possibly the axis* $\boldsymbol{\nu}$*.*

The result in Lemma 3.5.7 says that, in order for (3.32) to hold, the ellipsoid $\mathcal{E}^n$ has to have a special structure, namely it has to have equal radii for all axes except possibly the axis $\mathcal{L}_\mathcal{H}$ in (3.27). Mathematically, this special structure of the ellipsoid $\mathcal{E}^n$ can be represented as

$$\mathcal{E}^n \cap \mathcal{W} = \mathcal{B}^{n-1}(\mathbf{0}_n, \gamma), \tag{3.34}$$

where $\gamma \in (0, 1)$ is the radius of the $(n-1)$-dimensional ball $\mathcal{B}^{n-1}(\mathbf{0}_n, \gamma)$. Note that the fact that $\gamma \in (0, 1)$ follows from the structure of the original ellipsoid, i.e., $\left(K_\mathbf{N}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ since, by taking the eigendecomposition, we can write

$$\left(K_\mathbf{N}^{-1} + I_n\right)^{-\frac{1}{2}} = V(\Lambda^{-1} + I_n)^{-\frac{1}{2}} V^T,$$

which implies $\gamma < 1$ since all elements of $(\Lambda^{-1} + I_n)^{-\frac{1}{2}}$ are strictly smaller than one. We finally note that

$$\begin{aligned} \mathcal{E}^n \cap \mathcal{W} &\overset{(a)}{=} \mathrm{Proj}_\mathcal{W}(\mathcal{E}^n) \\ &\overset{(b)}{=} \mathrm{Proj}_\mathcal{W}\left(\left(K_\mathbf{N}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right), \end{aligned} \tag{3.35}$$

where the labeled equalities follow from: (a) the fact that $\mathcal{E}^n$ is a convex set and is symmetric with respect to $\mathcal{W}$; and (b) the projection property of Steiner symmetrization in Proposition 3.5.3. Thus, (3.34) becomes

$$\mathrm{Proj}_\mathcal{W}\left(\left(K_\mathbf{N}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right) = \mathcal{B}^{n-1}(\mathbf{0}_n, \gamma),$$

where $\gamma \in (0, 1)$. This concludes the proof of Lemma 3.5.6. □

3.5.3 Proof of the Implication $3) \Leftrightarrow 4)$

We prove that $3) \Leftrightarrow 4)$, namely we prove the next lemma.

Lemma 3.5.8. *Let* $\mathcal{Q} = \left\{ Q \in \mathcal{SO}(n) : \mathbf{q}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n \right\}$*, where* $\mathcal{SO}(n)$ *is the set of* $n \times n$ *real-valued orthonormal matrices, and* $\mathbf{q}_n$ *is the* n*-th column of* Q*. Then, a* $K_{\mathbf{N}}$ *is a solution for Lemma 3.5.6 if and only if*

$$\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} = Q \begin{bmatrix} \gamma I_{n-2} & 0_{n-2\times 2} \\ 0_{2\times n-2} & S \end{bmatrix} Q^T,$$

where $Q \in \mathcal{Q}$*,* $S = \left[\begin{smallmatrix} \gamma & v \\ v & a \end{smallmatrix}\right]$*, and* $\gamma \in (0,1)$*,* $a \in (0,1)$ *and* $v \in \mathbb{R}$ *satisfying* $v^2 < \min\{a\gamma, (1-a)(1-\gamma)\}$*.*

Proof. We start by noting that any n-dimensional ellipsoid can be represented in terms of a symmetric matrix. In particular, an n-dimensional ellipsoid defined as $K^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ with K being a positive definite matrix, can be equivalently represented as

$$K^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right) = \left\{ \mathbf{y} \in \mathbb{R}^n : \mathbf{y}^T K^{-1} \mathbf{y} \leq 1 \right\},$$

and hence

$$\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}}\mathcal{B}^n(\mathbf{0}_n, 1) = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \left(K_{\mathbf{N}}^{-1} + I_n\right) \mathbf{x} \leq 1 \right\}.$$

Now, let C be any $n \times (n-1)$ matrix whose columns form an orthonormal basis of the hyperplane $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T \mathbf{x} = 0\}$, which is an $(n-1)$-dimensional vector subspace of $\mathbb{R}^n$. Then, from [59], the relationship between the original ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$, which is specified by the matrix $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$, and its projection on the hyperplane $\mathcal{W}$ namely $\mathrm{Proj}_{\mathcal{W}}((K_{\mathbf{N}}^{-1} + I_n)^{-\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right))$, which is specified by B in the projection subspace, is given by the equation

$$B = C^T \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} C. \tag{3.36}$$

We want to find the sufficient and necessary conditions that ensure that the projection of the original ellipsoid $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n\left(\mathbf{0}_n, 1\right)$ onto the hyperplane $\mathcal{W}$ is an $(n-1)$-dimensional ball, i.e., in (3.36) we need $B = \gamma I_{n-1}$, where γ is the radius of the $(n-1)$-dimensional ball. We delegate the derivation of such necessary and sufficient conditions to Appendix A.7. □

3.5.4 Proof of the Implication $4) \Rightarrow 5)$

We here prove that $4) \Rightarrow 5)$, i.e., a $K_{\mathbf{N}}$ that satisfies Lemma 3.5.8 implies that $\mathcal{R}_{\tau,K_{\mathbf{N}}} = (K_{\mathbf{N}} + I_n)\,\mathcal{H}_\tau$, for all $\tau \in \mathcal{P}$. Towards this end, we leverage the following auxiliary lemma, the proof of which is in Appendix A.8.

Lemma 3.5.9. *Let $\tilde{\mathbf{Y}}_0 \sim \mathcal{N}\left(\mathbf{0}_n, \tilde{K}\right)$ with $\tilde{K} = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$ that satisfies the condition in Lemma 3.5.8. Then, there exists some $\beta \in (0,1)$ such that*

$$\Pr\left(\tilde{\mathbf{Y}}_0 \in \mathcal{H}_\tau\right) = \beta,\ \forall \tau \in \mathcal{P}. \tag{3.37}$$

Moreover, if $\tilde{\mathbf{y}} \in \mathcal{H}_\eta$, then

$$\Pr\left(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\eta\right) = \max_{\tau \in \mathcal{P}} \left\{\Pr(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\tau)\right\}. \tag{3.38}$$

We now leverage Lemma 3.5.9 to prove the implication $4) \Rightarrow 5)$, and hence to conclude the proof of Theorem 3.4.1. In particular, we have the following lemma.

Lemma 3.5.10. *Suppose that $K_{\mathbf{N}}$ satisfies the conditions in Lemma 3.5.8. Then,*

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = (K_{\mathbf{N}} + I_n)\,\mathcal{H}_\tau. \tag{3.39}$$

Proof. Let $\tilde{\mathbf{Y}} = \tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}}$ where $\tilde{\mathbf{Y}}_0 \sim \mathcal{N}\left(\mathbf{0}_n, \tilde{K}\right)$ with $\tilde{K} = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$, and $\tilde{\mathbf{y}} = (I_n + K_{\mathbf{N}})^{-1}\,\mathbf{y}$. Next, note that

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) &= \int_{\mathbf{x} \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} \\
&= \int_{\mathbf{x} \in \mathcal{H}_\tau} \frac{\mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\mathbf{x})^T K_{\mathbf{N}}^{-1}(\mathbf{y}-\mathbf{x})}}{\sqrt{(2\tau)^n \det(K_{\mathbf{N}})}} \frac{\mathrm{e}^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}}{\sqrt{(2\tau)^n}}\, \mathrm{d}\mathbf{x} \\
&= \int_{\mathbf{x} \in \mathcal{H}_\tau} \frac{\mathrm{e}^{-\frac{1}{2}\left(\mathbf{y}^T K_{\mathbf{N}}^{-1}\mathbf{y} - 2\mathbf{y}^T K_{\mathbf{N}}^{-1}\mathbf{x} + \mathbf{x}^T(K_{\mathbf{N}}^{-1} + I_n)\mathbf{x}\right)}}{(2\tau)^n \sqrt{\det(K_{\mathbf{N}})}}\, \mathrm{d}\mathbf{x} \\
&\overset{(a)}{=} C_{\mathbf{y}} \int_{\mathbf{x} \in \mathcal{H}_\tau} \frac{\mathrm{e}^{-\frac{1}{2}(\tilde{\mathbf{y}}-\mathbf{x})^T (K_{\mathbf{N}}^{-1} + I_n)(\tilde{\mathbf{y}}-\mathbf{x})}}{\sqrt{(2\tau)^n \det\left((K_{\mathbf{N}}^{-1} + I_n)^{-1}\right)}}\, \mathrm{d}\mathbf{x} \\
&\overset{(b)}{=} C_{\mathbf{y}} \Pr(\tilde{\mathbf{Y}} \in \mathcal{H}_\tau) \\
&= C_{\mathbf{y}} \Pr(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\tau),
\end{aligned} \tag{3.40}$$

where the labeled equalities follow from: (a) defining

$$C_{\mathbf{y}} = \frac{\sqrt{\det\left((K_{\mathbf{N}}^{-1} + I_n)^{-1}\right)}}{\sqrt{(2\tau)^n \det(K_{\mathbf{N}})}} \mathrm{e}^{-\frac{1}{2}\mathbf{y}^T K_{\mathbf{N}}^{-1}\mathbf{y} + \frac{1}{2}\tilde{\mathbf{y}}^T (K_{\mathbf{N}}^{-1} + I_n)\tilde{\mathbf{y}}};$$

and (b) noting that the integrand is equal to the multivariate Gaussian density $f_{\tilde{\mathbf{Y}}}(\cdot)$.

Now if $\tilde{\mathbf{y}} \in \mathcal{H}_\eta$ or equivalently if $\mathbf{y} \in (K_{\mathbf{N}} + I_n)\mathcal{H}_\eta$, in view of (3.40) and using Lemma 3.5.9, we have that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\eta) &= C_{\mathbf{y}} \Pr(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\eta) \\ &= C_{\mathbf{y}} \max_{\tau \in \mathcal{P}} \left\{ \Pr(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\tau) \right\} \\ &= \max_{\tau \in \mathcal{P}} \left\{ C_{\mathbf{y}} \Pr(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\tau) \right\} \\ &= \max_{\tau \in \mathcal{P}} \left\{ f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_\tau) \right\}. \end{aligned} \tag{3.41}$$

This indicates that $\mathcal{H}_\eta$ is an optimal decision for all $\mathbf{y} \in (K_{\mathbf{N}} + I_n)\mathcal{H}_\eta$. Thus, when $K_{\mathbf{N}}$ satisfies the conditions in Lemma 3.5.8, the optimal decision regions are given by

$$\mathcal{R}_{\tau, K_{\mathbf{N}}} = (K_{\mathbf{N}} + I_n)\mathcal{H}_\tau, \ \forall \tau \in \mathcal{P}. \tag{3.42}$$

This concludes the proof of Lemma 3.5.10, and of Theorem 3.4.1. □

3.6 Conclusion

We have considered a hypothesis testing framework to study a problem of data permutation recovery from an observation corrupted by correlated Gaussian noise. We have shown that the optimal decision regions may or may not be a linear transformation of the corresponding hypothesis regions depending on the noise covariance matrix. We have focused on the linear regime, which is appealing from a computational perspective as within it the optimal decoding is of polynomial complexity in the data size. We have characterized the optimal decision regions in the linear regime and shown that they are identical to the hypothesis of the observation multiplied by a permutation-independent linear function of the covariance matrix. We have discussed several practical implications of this result. For instance, we have shown that when the

data size is equal to two, the linear regime is the only regime, and when the data size is larger than two if the noise is memoryless then it must be isotropic to induce the linear regime. By leveraging the structure of the optimal decision regions, we have also derived the probability of error in terms of a volume of a region that consists of the intersection of a cone with a linear transformation of the unit radius ball.

Chapter 4

Probability of Error and Asymptotics

In this chapter, we focus on the probability of error in the permutation recovery problem.

4.1 Introduction

In this chapter, we study the data permutation recovery problem in the framework of an M-ary hypothesis testing defined in Chapter 2 with focus on the probability of error and its asymptotics. In particular, we consider a scenario where

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{4.1}$$

with $\mathbf{Y} \in \mathbb{R}^n$ being the n-dimensional noisy observation, $\mathbf{X} \in \mathbb{R}^n$ being the input data vector not necessarily isotropic Gaussian, and $\mathbf{N} \in \mathbb{R}^n$ being the noise vector distributed according to $\mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$. The goal is to estimate the ordering (that is, the permutation) of $\mathbf{X}$ based on the noisy observation of $\mathbf{Y}$. As a follow up to Chapter 3, where the permutation recovery problem was studied with a focus on characterizing the optimal decision criterion, this chapter studies the fundamental limits of this problem in terms of error probability under the constraint that a *linear decoder* (i.e., a linear estimator $A\mathbf{Y} + \mathbf{b}$ for some $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$ that are the same across all permutations, followed by a sorting operation) is employed. Studying the problem with such linear decoders is important for several reasons.

First, linear decoders can be optimal (i.e., they lead to the smallest probability of error) when the noise is isotropic, and the distribution of the input data vector is exchangeable; for

Table 4.1: Outline of our results under different data distributions.

Result \ **Data distribution**	Arbitrary	Exchangeable	i.i.d.
Error probability	Theorem 4.3.2	Corollary 4.3.3 & 4.3.6	–
Low-noise asymptotics	Theorem 4.4.1	Corollary 4.4.3	Proposition 4.4.4
High-noise asymptotics	Theorem 4.5.1& Proposition 4.5.3	Corollary 4.5.4	–
High-dimensional asymptotics	–	–	Theorem 4.6.1

this case, the optimal A and $\mathbf{b}$ are the identity matrix of dimension n and the all-zero vector of length n, respectively. Second, the optimal decoder can be linear even if the noise is colored; as shown in the previous chapter, for example, a linear decoder becomes optimal when the Gaussian noise covariance matrix $K_{\mathbf{N}}$ satisfies the geometrical condition that the projection of the n-dimensional ellipsoid induced by $K_{\mathbf{N}}$ onto the hyperplane $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^\top \mathbf{x} = 0\}$ is an $(n-1)$-dimensional ball; for this case, the optimal $\mathbf{b}$ is again the all-zero vector of dimension n, but the optimal A might not be anymore the identity matrix of dimension n. Third, linear decoders are suitable for practical implementations since they have at most a polynomial complexity in the data dimension n, while a naive optimal decoder (i.e., an exhaustive algorithm) has a factorial complexity in n. Finally, in scenarios where the statistics of the data distribution and/or of the noise are unknown, the linear decoder might be the only sensible choice.

Table 4.1 enables readers to easily access our main results, which are summarized as follows. Initially, we delineate the error probability associated with the data permutation recovery problem under the operation of a linear decoder. This characterization of the error probability remains valid across any continuous data and for Gaussian noise channel with arbitrary correlations amongst components in noise vector. Notably, we investigate the specific contexts of exchangeable data distributions and isotropic noise distributions, offering a thorough examination of their nuances.

Following this, in one hand, we explore the error probability behavior within the realm of minimal noise interference. For the purposes of our study, we adopt the premise of isotropic noise—characterized by a diagonal scalar covariance matrix, with σ representing the standard deviation of noise. Within this minimal noise environment (i.e., as $\sigma \to 0$), we observe that the error probability behaves linearly with σ. This increase manifests with a slope potentially

quadratic in relation to n, and correlates directly with the L_2 norm of the input data's probability distribution function. On the other hand, we extend our study to the high-noise scenario, analogous to our approach with the low-noise regime. Here, as $\sigma \to \infty$, we identify that the error probability can be expressed as a cumulative sum of anticipated spacings amongst the data order statistics. Notably, the expression $1 - 1/n!$ emerges as a predominant factor influencing the error probability across various data distributions of interest.

Concluding our analysis, we establish both upper and lower bounds for the probability of correctness, particularly when the input data vector, $\mathbf{X}$, comprises independently and identically distributed (i.i.d.) components. These bounds elucidate a decline in the probability of correctness to zero, exponentially as n extends towards infinity—characteristic of the high-dimensional domain. Moreover, we present a universal upper bound for the probability of correctness applicable to any sub-Gaussian i.i.d. data distributions, alongside more stringent bounds for scenarios wherein $\mathbf{X}$ adheres to an i.i.d. Gaussian model.

4.2 Preliminaries: Generalized Spacing

The data permutation recovery problem is naturally related to the field of order statistics [53]. In particular, as it will become clear throughout the paper, the spacing [60] between order statistics will play a significant role. We here formally introduce the concepts of spacing and generalized spacing using the order statistics of $\mathbf{X}$. The i-th order statistics of $\mathbf{X} \in \mathbb{R}^n$ (denoted by $X_{i:n}$) is a random variable that follows the distribution of the i-th smallest value among the n entries of $\mathbf{X}$. For instance, $X_{1:n}$ (respectively, $X_{n:n}$) denotes the random variable with the distribution of the smallest (respectively, largest) value of $\mathbf{X}$. The i-th spacing (denoted by W_i) is then formally defined as follows.

Definition 4.2.1 (Spacing [60])**.** Given a random vector $\mathbf{X} \in \mathbb{R}^n$, its i-th spacing is defined as

$$W_i = X_{i+1:n} - X_{i:n},\ \forall i \in [1 : n-1], \tag{4.2}$$

where $X_{i:n}$ is the i-th order statistics of $\mathbf{X}$.

By stacking together W_i for all $i \in [1 : n-1]$, we obtain $\mathbf{W} \in \mathbb{R}^{n-1}$, which represents the gap between the order statistics [53] of $\mathbf{X}$ (i.e., the sorted version of $\mathbf{X}$). Since $\mathbf{W}$ measures the absolute difference between nearest values of $\mathbf{X}$, the support of $\mathbf{W}$ is the non-negative space,

i.e., $\Pr(\mathbf{W} \geq \mathbf{0}_n) = 1$. When $\mathbf{X}$ is exchangeable[1], it will be convenient to represent the spacing of $\mathbf{X}$ in the following conditional form[2],

$$\mathbf{W} \stackrel{d}{=} T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau, \ \forall \tau \in \mathcal{P}, \tag{4.3}$$

where $T_\tau \in \mathbb{R}^{(n-1)\times n}$ is given by

$$(T_\tau)_{i,j} = \mathbb{1}\{j = \tau_{i+1}\} - \mathbb{1}\{j = \tau_i\}, \tag{4.4}$$

with $\mathbb{1}_{\{x=y\}} = 1$ if and only if $x = y$ and equal to zero otherwise. For instance, let $n = 4$ and consider $\tau = (4, 2, 1, 3)$; then, we have that

$$T_{(4,2,1,3)} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}. \tag{4.5}$$

In particular, (4.3) follows from the fact that, by letting $\eta = (1, 2, \ldots, n)$, the spacing can be expressed as

$$\begin{aligned} \mathbf{W} &\stackrel{(a)}{=} T_\eta \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\eta \\ &\stackrel{(b)}{=} T_\tau P \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\eta \\ &\stackrel{(c)}{=} T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau, \end{aligned} \tag{4.6}$$

where the labeled equalities follow from: (a) Definition 4.2.1; (b) letting $T_\eta = T_\tau P$, where P is the permutation matrix that permutes $\mathbf{x} \in \mathcal{H}_\eta$ to $P\mathbf{x} \in \mathcal{H}_\tau$; and (c) the fact that $\mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\eta \stackrel{d}{=} P^\top \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau$ due to the exchangeability and $PP^\top = I_n$.

The expression of $\mathbf{W}$ in (4.3) holds under the assumption that $\mathbf{X}$ is exchangeable. In order to analyze the case of an arbitrary (i.e., not necessarily exchangeable) distribution on $\mathbf{X}$, we will need the concept of τ-spacing of $\mathbf{X}$ for the permutation $\tau \in \mathcal{P}$, which we next formally define.

Definition 4.2.2 (τ-spacing)**.** Given a random vector $\mathbf{X} \in \mathbb{R}^n$, the τ-spacing of $\mathbf{X}$ for $\tau \in \mathcal{P}$ is

[1] A sequence of random variables $U_1, \ldots, U_n$ is said to be exchangeable if, for any permutation $(\tau_1, \ldots, \tau_n)$ of the indices $[1 : n]$, we have that $(U_1, \ldots, U_n) \stackrel{d}{=} (U_{\tau_1}, \ldots, U_{\tau_n})$.

[2] With reference to (4.3), $T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau$ represents the random vector $T_\tau \mathbf{X}$ conditioned on the event $\mathbf{X} \in \mathcal{H}_\tau$.

defined as

$$\mathbf{W}_\tau = T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau, \tag{4.7}$$

where T_τ is given in (5.15). In other words, $\mathbf{W}_\tau$ is the spacing of $\mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau$.

We note that, when $\mathbf{X} \in \mathbb{R}^n$ is exchangeable, then $\mathbf{W} \stackrel{d}{=} \mathbf{W}_\tau$ for all $\tau \in \mathcal{P}$, where $\mathbf{W}$ and $\mathbf{W}_\tau$ are defined in Definition 4.2.1 and Definition 4.2.2, respectively. However, in general $\mathbf{W} \stackrel{d}{\neq} \mathbf{W}_\tau$.

4.3 Probability of Error with Linear Decoder

In this section, we focus on characterizing the probability of error of the data permutation recovery problem described in Section 2.2. Given the hypothesis and decision regions defined in (2.1) and (2.4), respectively, we have that the *optimal* error probability P_e is given by

$$P_e = 1 - P_c, \tag{4.8a}$$

$$P_c = \sum_{\tau \in \mathcal{P}} \Pr\left(\{\mathbf{Y} \in \mathcal{R}_{\tau, K_\mathbf{N}}\} \cap \{\mathbf{X} \in \mathcal{H}_\tau\}\right), \tag{4.8b}$$

where P_c is the probability of correctness. Note that the error probability in (4.8) is the minimum error probability as the decision regions $\mathcal{R}_{\tau,K_\mathbf{N}}$'s in (2.4) follow the MAP criterion. Our focus is on assessing and analyzing the probability of error when a *linear* decoder is employed. The linear decoder $\phi_{\text{lin}} : \mathbb{R}^n \to \mathcal{P}$, based on the parameters $\tilde{A}$ and $\tilde{\mathbf{b}}$, is defined as follows.

Definition 4.3.1 (Linear decoder)**.** The linear decoder parameterized by $\tilde{A} \in \mathbb{R}^{n \times n}$ and $\tilde{\mathbf{b}} \in \mathbb{R}^n$ is defined such that, for all $\mathbf{y} \in \mathbb{R}^n$, we have

$$\phi_{\text{lin}}(\mathbf{y}; \tilde{A}, \tilde{\mathbf{b}}) = \pi_{\tilde{A}\mathbf{y}+\tilde{\mathbf{b}}}, \tag{4.9}$$

where $\pi_{\tilde{A}\mathbf{y}+\tilde{\mathbf{b}}}$ is the permutation according to which the vector $\tilde{A}\mathbf{y} + \tilde{\mathbf{b}}$ is sorted.

The linear decoder in Definition 4.3.1 with parameters $\tilde{A}$ and $\tilde{\mathbf{b}}$ first computes a permutation-independent linear transformation $\mathbf{y}_\ell$ of $\mathbf{y}$, i.e., $\mathbf{y}_\ell = \tilde{A}\mathbf{y} + \tilde{\mathbf{b}}$, where $\tilde{A} \in \mathbb{R}^{n \times n}$ and $\tilde{\mathbf{b}} \in \mathbb{R}^n$ are the same for all permutations, and then it outputs the permutation of $\mathbf{y}_\ell$. The corresponding

decision regions in (2.4) when the linear decoder is used become

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = A\mathcal{H}_\tau + \mathbf{b},\ \forall \tau \in \mathcal{P}. \tag{4.10}$$

Note that the parameters $(A, \mathbf{b})$ in (4.10) and $(\tilde{A}, \tilde{\mathbf{b}})$ in (4.9) are related as $A = \tilde{A}^{-1}$ and $\mathbf{b} = -A^{-1}\tilde{\mathbf{b}}$. Our choice of assessing the probability of error performance when a linear decoder is employed stems primarily from three factors. First, a linear decoder has low complexity (at most polynomial in the data-vector dimension n) compared to a brute force evaluation of the optimal test in (2.4), which has a practically prohibitive complexity of $n!$. Second, for some cases, such as when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$, it has been shown in Chapter 3 that a linear decoder can indeed be optimal, i.e., it minimizes the probability of error, under certain conditions on the noise covariance matrix $K_{\mathbf{N}}$. Third, even when the conditions in Theorem 3.4.1 are not satisfied, extensive numerical evaluations have shown that the probability of error incurred by using a linear decoder ϕ_{lin} parameterized by $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$ is very close to the one incurred by the optimal MAP decoder ϕ_{MAP}. For instance, Fig. 4.1 (obtained by using a Monte Carlo simulation with 10^5 iterations) considers the case $n = 3$ and shows that ϕ_{lin} (red curves) performs closely to ϕ_{MAP} (blue curves) for the case when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_3, K_{\mathbf{X}})$ with $K_{\mathbf{X}} \in \{K_1, K_2\}$ where

$$K_1 = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2} & 1 \end{bmatrix} \text{ and } K_2 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 1 \end{bmatrix}. \tag{4.11}$$

It is worth noting that K_1 and K_2 above do not satisfy the optimality conditions in Theorem 3.4.1 since $\mathbf{X}$ is not exchangeable.

4.3.1 Probability of Error

We now derive an expression for the probability of error when a linear decoder is used, which is denoted by $P_{e,\text{lin}}$ and is given by the theorem below.

Theorem 4.3.2. *Let $\mathbf{X} \in \mathbb{R}^n$ be a continuous random vector, and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$. Then, for any invertible A and $\mathbf{b}$ defined in* (4.10) *and any noise covariance matrix $K_{\mathbf{N}}$, the probability*

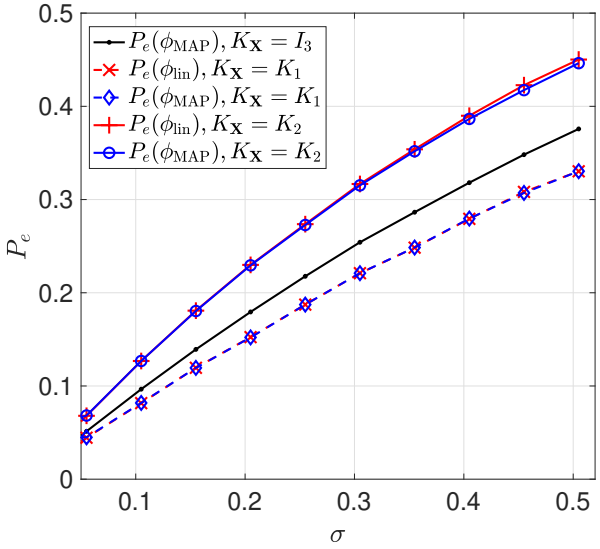


Figure 4.1: Comparison of P_e using the decoders ϕ_{MAP} (blue curves) and ϕ_{lin} (red curves), where ϕ_{lin} uses $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$. We set $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_3, K_{\mathbf{X}})$ with $K_{\mathbf{X}} \in \{I_3, K_1, K_2\}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_3, \sigma^2 I_3)$, where K_1 and K_2 are given in (4.11).

of error when the linear decoder is employed is given by

$$P_{e,\text{lin}} = 1 - \sum_{\tau \in \mathcal{P}} \mathbb{E}\Big[Q_{\tilde{K}_\tau}\big(-T_\tau A^{-1}(\mathbf{X} - \mathbf{b})\big) \,\Big|\, \mathbf{X} \in \mathcal{H}_\tau \Big] P_{\mathbf{X}}(\mathcal{H}_\tau), \tag{4.12}$$

where $\tilde{K}_\tau = T_\tau A^{-1} K_{\mathbf{N}} A^{-\top} T_\tau^\top \in \mathbb{R}^{(n-1)\times(n-1)}$ *with* $T_\tau, \tau \in \mathcal{P}$ *given by* (4.4)*, and where* $Q_{\tilde{K}_\tau}(\cdot)$ *is the multivariate Gaussian Q-function with covariance matrix* $\tilde{K}_\tau$*.*

Proof. By substituting the decision regions in (4.10) inside the probability of correctness in (4.8) and by using the Bayes' theorem, we obtain

$$\begin{aligned} P_{c,\text{lin}} &= \sum_{\tau \in \mathcal{P}} \Pr\left(\mathbf{Y} \in A\mathcal{H}_\tau + \mathbf{b}, \mathbf{X} \in \mathcal{H}_\tau\right) \\ &= \sum_{\tau \in \mathcal{P}} \Pr\left(\mathbf{Y} \in A\mathcal{H}_\tau + \mathbf{b} \mid \mathbf{X} \in \mathcal{H}_\tau\right) P_{\mathbf{X}}(\mathcal{H}_\tau) \\ &= \sum_{\tau \in \mathcal{P}} \left[\mathbb{E}\left[\Pr\left(\mathbf{X} + K_{\mathbf{N}}^{\frac{1}{2}} \mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\tau \,\middle|\, \mathbf{X} \right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau \right] \times P_{\mathbf{X}}(\mathcal{H}_\tau) \right], \end{aligned} \tag{4.13}$$

where the last equality follows by the law of total expectation and by the fact that $\mathbf{Y} = \mathbf{X} +$

$K_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z}$ with $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$.

We now focus on the conditional probability inside the conditional expectation in (4.13). For each $\mathcal{H}_\tau,\ \forall \tau \in \mathcal{P}$ we have that

$$\begin{aligned}\Pr\left(\mathbf{X} + K_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\tau \;\middle|\; \mathbf{X}\right) &= \Pr\left(A^{-1}(\mathbf{X}-\mathbf{b}) + A^{-1}K_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z} \in \mathcal{H}_\tau \;\middle|\; \mathbf{X}\right) \\ &= \Pr\left(A^{-1}(\mathbf{X}-\mathbf{b}) + \mathbf{U} \in \mathcal{H}_\tau \mid \mathbf{X}\right), \end{aligned} \tag{4.14}$$

where the last equality follows by letting $\mathbf{U} = A^{-1}K_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z}$. Note that $\mathbf{U} \sim \mathcal{N}(\mathbf{0}_n, A^{-1}K_{\mathbf{N}}A^{-\top})$. Then, given $\mathbf{X}$, the event inside the conditional probability in (4.14) can be expressed as

$$\begin{aligned}\left\{A^{-1}(\mathbf{X}-\mathbf{b}) + \mathbf{U} \in \mathcal{H}_\tau\right\} &= \bigcap_{k=1}^{n-1}\left\{\left(A^{-1}(\mathbf{X}-\mathbf{b})\right)_{\tau_k} + U_{\tau_k} \le \left(A^{-1}(\mathbf{X}-\mathbf{b})\right)_{\tau_{k+1}} + U_{\tau_{k+1}}\right\} \\ &= \bigcap_{k=1}^{n-1}\left\{\left(A^{-1}(\mathbf{X}-\mathbf{b})\right)_{\tau_k} - \left(A^{-1}(\mathbf{X}-\mathbf{b})\right)_{\tau_{k+1}} \le U_{\tau_{k+1}} - U_{\tau_k}\right\} \\ &= \left\{-T_\tau A^{-1}(\mathbf{X}-\mathbf{b}) \le T_\tau \mathbf{U}\right\}, \end{aligned} \tag{4.15}$$

where $(A^{-1}(\mathbf{X}-\mathbf{b}))_{\tau_k}$ denotes the τ_k-th entry of $A^{-1}(\mathbf{X}-\mathbf{b})$, and the last equality follows by using the definition of $T_\tau, \tau \in \mathcal{P}$ in (5.15). By introducing a random vector $\mathbf{V}_\tau = T_\tau \mathbf{U} \sim \mathcal{N}(\mathbf{0}_{n-1}, \tilde{K}_\tau)$, where $\tilde{K}_\tau = T_\tau A^{-1}K_{\mathbf{N}}A^{-\top}T_\tau^\top$, and by leveraging (4.15), we obtain an equivalent expression for (4.14) as

$$\Pr\left(\mathbf{X} + K_{\mathbf{N}}^{\frac{1}{2}}\mathbf{Z} - \mathbf{b} \in A\mathcal{H}_\tau \;\middle|\; \mathbf{X}\right) = \Pr\left(-T_\tau A^{-1}(\mathbf{X}-\mathbf{b}) \le \mathbf{V}_\tau \mid \mathbf{X}\right). \tag{4.16}$$

By substituting this into (4.13), we obtain

$$\begin{aligned}P_{c,\text{lin}} &= \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[\Pr\left(-T_\tau A^{-1}(\mathbf{X}-\mathbf{b}) \le \mathbf{V}_\tau \,\middle|\, \mathbf{X}\right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right] P_{\mathbf{X}}(\mathcal{H}_\tau) \\ &= \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[Q_{\tilde{K}_\tau}\left(-T_\tau A^{-1}(\mathbf{X}-\mathbf{b})\right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right] P_{\mathbf{X}}(\mathcal{H}_\tau), \end{aligned} \tag{4.17}$$

where the last equality follows by letting $Q_{\tilde{K}_\tau}(\cdot)$ be the multivariate Gaussian Q-function with covariance matrix $\tilde{K}_\tau$. We conclude the proof of Theorem 4.3.2 by using $P_{e,\text{lin}} = 1 - P_{c,\text{lin}}$. $\square$

Next, we evaluate the probability of error expression in Theorem 4.3.2 for two practically

relevant cases, in which the expression for $P_{e,\text{lin}}$ is also slightly simpler.

4.3.2 Exchangeable Data Distribution

We here investigate the effect of data exchangeability. Let us assume that $\mathbf{X}$ is exchangeable. By the definition of exchangeability, we have that $\mathbf{X} \overset{d}{=} P\mathbf{X}$ for any permutation matrix P. Thus, for any $\tau \in \mathcal{P}$ and $\eta \in \mathcal{P}$, we have that

$$P_{\mathbf{X}}(\mathcal{H}_\tau) = P_{\mathbf{X}}(\mathcal{H}_\eta), \tag{4.18}$$

which results in

$$\sum_{\tau \in \mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau) = 1 \implies P_{\mathbf{X}}(\mathcal{H}_\tau) = \frac{1}{n!},\ \forall \tau \in \mathcal{P}. \tag{4.19}$$

By substituting (4.19) inside (4.12) in Theorem 4.3.2, we obtain the following corollary that provides the probability of error under the exchangeability assumption on $\mathbf{X}$.

Corollary 4.3.3. *Let $\mathbf{X} \in \mathbb{R}^n$ be an exchangeable random vector, and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_{\mathbf{N}})$. Then, the probability of error expression in Theorem 4.3.2 reduces to*

$$P_{e,\text{lin}} = 1 - \frac{1}{n!} \sum_{\tau \in \mathcal{P}} \mathbb{E}\left[Q_{\tilde{K}_\tau}\left(-T_\tau A^{-1}(\mathbf{X} - \mathbf{b}) \right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau \right]. \tag{4.20}$$

Remark 4.3.4. The exchangeability assumption in Corollary 4.3.3 can be considered as mild. Exchangeable, in fact, includes data that does not need to be necessarily i.i.d., but can be correlated. For instance, any convex combination of i.i.d. random variables and any spherically contoured distribution are exchangeable. Moreover, assuming an exchangeable data distribution is reasonable whenever the data has no natural order, e.g., relational data such as social network users, ratings, and preference data [61].

4.3.3 Exchangeable Data Distribution and Isotropic Noise

We here consider the case when $\mathbf{X}$ is exchangeable and the noise is isotropic, i.e., the covariance matrix of $\mathbf{N}$ is a diagonal matrix with constant entries, i.e., $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. These assumptions substantially simplify the result and make it easy to decode the permutation. For

instance, we have that the optimal decision regions are given by $\mathcal{R}_{\tau,\sigma} = \mathcal{H}_\tau, \tau \in \mathcal{P}$ as shown in the following lemma (proved in Appendix B.1), which is a generalization of [1, Theorem 1].

Lemma 4.3.5. *Assume that $\mathbf{X} \in \mathbb{R}^n$ is exchangeable and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Then, given an observation $\mathbf{y}$ of $\mathbf{Y}$, we have that*

$$\phi_{\text{opt}}(\mathbf{y}) = \phi_{\text{lin}}(\mathbf{y}; cI_n, \mathbf{0}_n), \tag{4.21}$$

for any $c > 0$. As a result, $P_e = P_{e,\text{lin}}$.

With reference to (4.9) and (4.10), by setting $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$, Lemma 4.3.5 states that the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is optimal regardless of the value of σ. By substituting these values inside $\tilde{K}_\tau \in \mathbb{R}^{(n-1)\times(n-1)}$ in Theorem 4.3.2, we obtain

$$\begin{aligned} \tilde{K}_\tau &= T_\tau A^{-1} K_{\mathbf{N}} A^{-\top} T_\tau^\top = \sigma^2 T_\tau T_\tau^\top = \sigma^2 \tilde{K}, \\ (\tilde{K})_{i,j} &= \begin{cases} 2 & i = j, \\ -1 & i = j+1 \text{ and } j = i+1, \\ 0 & \text{otherwise}, \end{cases} \end{aligned} \tag{4.22}$$

that is, $\tilde{K} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a tridiagonal Toeplitz matrix. Thus, with $\tilde{K}_\tau = \sigma^2 \tilde{K}$ for all $\tau \in \mathcal{P}$, the probability of error expression in Theorem 4.3.2 (or Corollary 4.3.3) reduces to

$$P_{e,\text{lin}} = 1 - \frac{1}{n!} \sum_{\tau \in \mathcal{P}} \mathbb{E}\left[Q_{\sigma^2 \tilde{K}}\left(-T_\tau \mathbf{X}\right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau \right]. \tag{4.23}$$

We note that $\sigma^2 \tilde{K}$ is independent of $\tau \in \mathcal{P}$, and $\mathbf{W} \overset{d}{=} T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau,\ \forall \tau \in \mathcal{P}$ as stated in (4.3). Hence, the conditional expectation in (4.23) is constant in $\tau \in \mathcal{P}$ and can be written as

$$P_{e,\text{lin}} = 1 - \frac{1}{n!} \mathbb{E}\left[Q_{\sigma^2 \tilde{K}}\left(-\mathbf{W}\right) \right] \sum_{\tau \in \mathcal{P}} 1 = 1 - \mathbb{E}\left[Q_{\sigma^2 \tilde{K}}\left(-\mathbf{W}\right) \right],$$

which gives the minimum probability of error in the isotropic noise scenario with exchangeable $\mathbf{X} \in \mathbb{R}^n$. The following corollary states this result formally.

Corollary 4.3.6. *Let $\mathbf{X} \in \mathbb{R}^n$ be an exchangeable random vector and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Let*

$\mathbf{W}$ *be the spacing of* $\mathbf{X}$*. Then, the minimum probability of error is given by* $P_e = P_{e,\mathrm{lin}}$ *with*

$$P_{e,\mathrm{lin}} = 1 - \mathbb{E}\left[Q_{\sigma^2\tilde{K}}\left(-\mathbf{W}\right)\right], \tag{4.24}$$

where $\tilde{K}$ *is defined in* (4.22) *and where* $Q_{\sigma^2\tilde{K}}(\cdot)$ *is the multivariate Gaussian Q-function with covariance matrix* $\sigma^2\tilde{K}$*.*

With the goal to better understand the behavior of the probability of error, in the remaining of this chapter, we will focus on the isotropic noise scenario, i.e., we will assume that $K_{\mathbf{N}} = \sigma^2 I_n$. As will be argued below, for many cases of interest, this assumption is without loss of generality.

4.4 Low-noise Regime

In this section, we study the rate of convergence of $P_{e,\mathrm{lin}}$ in the low-noise regime (i.e., $\sigma \to 0$). In particular, we consider the error probability of the data permutation recovery problem when the noise is isotropic, i.e., $K_{\mathbf{N}} = \sigma^2 I_n$. Under this assumption, the regions $\mathcal{R}_{\tau,K_{\mathbf{N}}}$, $\tau \in \mathcal{P}$ in (4.10) depend on $K_{\mathbf{N}}$ only through σ, and hence, we let $\mathcal{R}_{\tau,K_{\mathbf{N}}} = \mathcal{R}_{\tau,\sigma}$.

4.4.1 Arbitrary Data Distribution

We here focus on the asymptotic behavior of the probability of error in Theorem 4.3.2 in the low-noise regime, without any assumption on the distribution of $\mathbf{X}$. We start by noting that we expect the probability of error to be close to zero when $\sigma \to 0$. However, when a linear decoder $\phi_{\mathrm{lin}}(\cdot; A, \mathbf{b})$ is used, the probability of error might not be close to zero if A and $\mathbf{b}$ are not properly chosen. To see this, consider the extreme case $\sigma = 0$ (noiseless case), and consider the linear decoder $\phi_{\mathrm{lin}}(\cdot; P, \mathbf{0}_n)$, where P is any permutation matrix except for the identity matrix I_n. Since $\sigma = 0$, we observe $\mathbf{y} = \mathbf{x}$; however, $\phi_{\mathrm{lin}}(\cdot; P, \mathbf{0}_n)$ declares the permutation of $P\mathbf{y}$ as the estimated permutation of $\mathbf{x}$, which incurs a probability of error equal to one. Given this, in order to avoid a non-zero error probability when $\sigma \to 0$, we set the linear decoder to be $\phi_{\mathrm{lin}}(\cdot; I_n, \mathbf{0}_n)$. As argued above, this linear decoder gives the correct value of $P_{e,\mathrm{lin}}$ for $\sigma = 0$ for any distribution on $\mathbf{X}$ (see also Fig. 4.1 for a simulation result that shows its closeness to the optimal MAP decoder).

With the linear decoder $\phi_{\text{lin}}(\cdot, I_n, \mathbf{0}_n)$, we are now interested in characterizing $\lim_{\sigma\to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma}$ as an asymptotic behavior of $P_{e,\text{lin}}(\sigma)$ in the low-noise regime (we use $P_{e,\text{lin}}(\sigma)$ to highlight the fact that the probability of error is a function of σ). The next result, the proof of which can be found in Appendix B.2, shows the behavior of $P_{e,\text{lin}}(\sigma)$ in the low-noise regime for an arbitrary distribution of $\mathbf{X}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$.

Theorem 4.4.1. *Let $\mathbf{X} \in \mathbb{R}^n$ be a continuous random vector and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume that the τ-spacing of $\mathbf{X}$ in* (4.7)*, namely $\mathbf{W}_\tau$, is such that, for any $\tau \in \mathcal{P}$,*

$$\max_w f_{(\mathbf{W}_\tau)_i}(w) < \infty, \ \forall i, \tag{4.25a}$$

$$\max_{u,v} f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(u, v) < \infty, \ \forall s, t. \tag{4.25b}$$

Then, by using the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$, we have that

$$P_{e,\text{lin}}(\sigma) = \sum_{\tau\in\mathcal{P}} \sum_{i=1}^{n-1} P_{\mathbf{X}}(\mathcal{H}_\tau) \frac{f_{(\mathbf{W}_\tau)_i}(0^+)}{\sqrt{\pi}} \sigma + O(\sigma^2), \tag{4.26}$$

where $f_{(\mathbf{W}_\tau)_i}(0^+) = \lim_{w\to 0^+} f_{(\mathbf{W}_\tau)_i}(w)$. Consequently, in the low-noise regime, we have that

$$\lim_{\sigma\to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \sum_{\tau\in\mathcal{P}} \sum_{i=1}^{n-1} P_{\mathbf{X}}(\mathcal{H}_\tau) \frac{f_{(\mathbf{W}_\tau)_i}(0^+)}{\sqrt{\pi}}. \tag{4.27}$$

Theorem 4.4.1 provides the exact rate of convergence of $P_{e,\text{lin}}(\sigma)$ in the low-noise regime and the first-order approximation of $P_{e,\text{lin}}(\sigma)$ for a large class of distributions of $\mathbf{X}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$ when $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is used. Theorem 4.4.1 demonstrates that in such a setting and provided that (4.25) holds, the probability of error is a linear function of σ in the low-noise regime, and its slope is determined by the density function of the τ-spacing of $\mathbf{X}$ for all $\tau \in \mathcal{P}$.

Remark 4.4.2. Theorem 4.4.1 holds for any distribution of $\mathbf{X}$ under the assumption in (4.25) and hence, the assumption $K_{\mathbf{N}} = \sigma^2 I_n$ is without loss of generality. In particular, the result for $\mathbf{X} \sim f_{\mathbf{X}}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 K)$ with positive definite $K \in \mathbb{R}^{n\times n}$ is the same as the result for $\mathbf{X} \sim f_{K^{-\frac{1}{2}}\mathbf{X}}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$.

Next, we analyze Theorem 4.4.1 for two cases for which the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is optimal, i.e., $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$.

4.4.2 Exchangeable Data Distribution

We here consider a distribution of $\mathbf{X}$ that is exchangeable, which implies that $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is optimal (see Lemma 4.3.5) with the corresponding error probability given in Corollary 4.3.6. Since for an exchangeable $\mathbf{X}$, we have that $\mathbf{W} \overset{d}{=} \mathbf{W}_\tau,\ \forall \tau \in \mathcal{P}$, instead of the τ-spacing in Definition 4.2.2, we make use of the spacing of $\mathbf{X}$ in Definition 4.2.1. With this, we obtain the next corollary, which shows the low-noise asymptotic of the minimum $P_e(\sigma)$ under the exchangeability assumption.

Corollary 4.4.3. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume that the spacing of $\mathbf{X}$ in* (4.3)*, namely $\mathbf{W}$, is such that*

$$\max_w f_{W_i}(w) < \infty,\ \forall i \tag{4.28a}$$

$$\max_{u,v} f_{W_s,W_t}(u,v) < \infty,\ \forall s,t. \tag{4.28b}$$

Then, by using the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$, we have that $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$ with

$$P_{e,\text{lin}}(\sigma) = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}} \sigma + O(\sigma^2), \tag{4.29}$$

where $f_{W_i}(0^+) = \lim_{w \to 0^+} f_{W_i}(w)$. Consequently, in the low-noise regime, we have that

$$\lim_{\sigma \to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}. \tag{4.30}$$

4.4.3 i.i.d. Data Distribution

As shown next in the case of i.i.d. data (which is a special case of exchangeable data), the low-noise expansion of $P_{e,\text{lin}}(\sigma)$ has an explicit dependence on the dimension of the data. In particular, the following proposition (proved in Appendix B.3) shows that the asymptotic behavior of $P_e(\sigma)$ in the low-noise regime can be determined by the L_2-norm of the pdf of any single entry of $\mathbf{X}$ and has a quadratic dependence on the dimension of $\mathbf{X}$.

Proposition 4.4.4. *Let $\mathbf{X} \in \mathbb{R}^n$ consist of n i.i.d. random variables generated according to X, and assume that*

$$\|f_X\|_2 < \infty, \tag{4.31}$$

where $\|f_X\|_2 = \sqrt{\int_{-\infty}^{\infty} f_X^2(x)\,\mathrm{d}x}$. *Then, we have that* $P_e(\sigma) = P_{e,\mathrm{lin}}(\sigma)$ *with*

$$P_{e,\mathrm{lin}}(\sigma) = \frac{n(n-1)}{\sqrt{\pi}}\|f_X\|_2^2\sigma + O(\sigma^2). \tag{4.32}$$

Consequently, in the low-noise regime, we have that

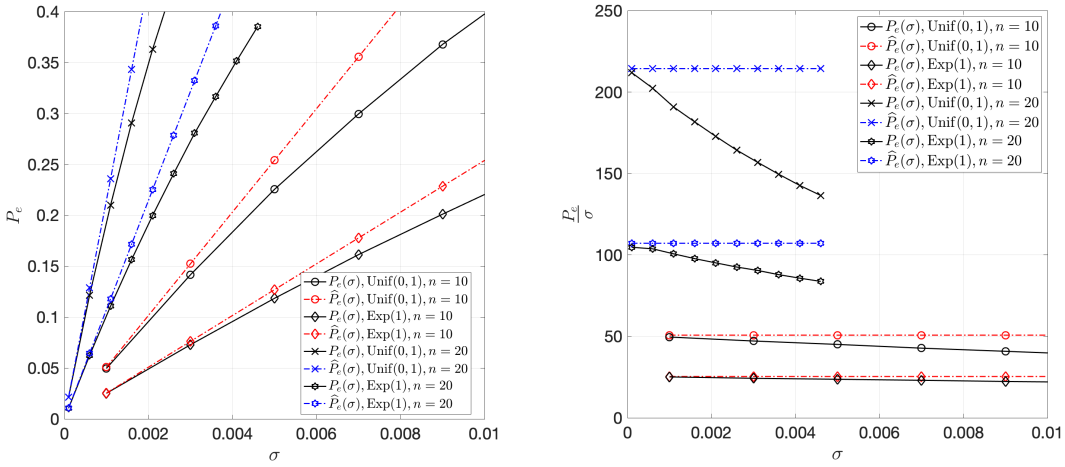
$$\lim_{\sigma\to 0}\frac{P_{e,\mathrm{lin}}(\sigma)}{\sigma} = \frac{n(n-1)}{\sqrt{\pi}}\|f_X\|_2^2. \tag{4.33}$$

From Proposition 4.4.4, we observe that the i.i.d. assumption considerably simplifies the expression for the asymptotic of $P_e(\sigma)$ when $\sigma \to 0$, i.e., we do not need to know f_{W_i} (see Corollary 4.4.3) to evaluate the rate of convergence. Moreover, Proposition 4.4.4 shows that in the low-noise regime the rate of convergence only depends on the L_2-norm of the pdf of X, hence implying that the i.i.d. data distribution assumption weakens the data dependence on the probability of error. In other words, the exact distribution of X is not required to compute the rate of convergence of $P_e(\sigma)$ and its first-order approximation. It is also worth noting that the fact that the error probability grows proportionally to $\|f_X\|_2^2$ is reasonable. This is because the spacing between the coordinates of $\mathbf{X}$ decreases as $\|f_X\|_2^2$ increases; for instance, for the location-scale family of probability distributions on $\mathbf{X}$, it is not difficult to see that $\frac{1}{\|f_X\|_2^2}$ is proportional to the scale parameter of the family that is proportional to the spacing. Beyond the Gaussian noise, the low-noise asymptotic of $P_e(\sigma)$ under more broader noise distributions is studied in Chapter 5.

We conclude this section with a couple of remarks regarding the results on the low-noise asymptotics, and with the evaluation of (4.33) for a few distributions.

Remark 4.4.5. Theorem 4.4.1, Corollary 4.4.3, and Proposition 4.4.4 show that, in the low-noise regime, $P_e(\sigma)$ grows linearly with σ, as we also empirically observe from Fig. 5.2 (solid curves). These results also allow us to obtain the first-order approximation of $P_e(\sigma)$, which is also shown in Fig. 4.2 (dashed curves). The approximation $\widehat{P}_e(\sigma)$ is close to $P_e(\sigma)$ when σ is small, as shown in Fig. 4.2. Lastly, we can conclude that, in the low-noise regime, $P_e(\sigma)$ can quadratically increase with n, i.e., the error probability is highly sensitive to σ when n is large.

Remark 4.4.6. An example of a pdf that does not satisfy the condition in Proposition 4.4.4 is $f_X(x) = \frac{1}{2\sqrt{x}}, x \in [0,1]$. Note that the spacing of such a distribution heavily concentrates around zero, and $P_e(\sigma)$ may no longer be linear in σ. An interesting future direction would be



(a) P_e v.s. σ (b) $\frac{P_e}{\sigma}$ v.s. σ

Figure 4.2: Comparison between $P_e(\sigma)$ (solid curves) and its first-order approximation $\widehat{P}_e(\sigma)$ (dashed curves). We set $X \sim \text{Unif}(0,1)$ and $X \sim \text{Exp}(1)$ with dimension $n \in \{10, 20\}$.

to characterize the low-noise asymptotics also for distributions with $\|f_X\|_2 = \infty$.

Example 4.4.7. Consider $X \sim \text{Unif}(a,b),\ 0 \le a < b < \infty$. Then, $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$ with

$$\lim_{\sigma \to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \frac{n(n-1)}{(b-a)\sqrt{\pi}}.$$

Example 4.4.8. Consider $X \sim \text{Exp}(\lambda),\ \lambda > 0$. Then, $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$ with

$$\lim_{\sigma \to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \frac{\lambda n(n-1)}{2\sqrt{\pi}}.$$

Example 4.4.9. Consider $X \sim \mathcal{N}(0,1)$. Then, $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$ with

$$\lim_{\sigma \to 0} \frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \frac{n(n-1)}{2\pi}.$$

4.5 High-noise Regime

In this section, we study the rate of convergence of $P_{e,\text{lin}}(\sigma)$ in the high-noise regime (i.e., $\sigma \to \infty$). As in Section 4.4 for the low-noise regime, we assume that the noise is isotropic, i.e.,

$K_{\mathbf{N}} = \sigma^2 I_n$. It is worth noting that in the high-noise regime when $\sigma \to \infty$, an optimal decoder would incur a probability of error equal to

$$\lim_{\sigma \to \infty} P_e(\sigma) = 1 - \max_{\tau} P_{\mathbf{X}}(\mathcal{H}_\tau) = P_e(\infty). \tag{4.34}$$

The result above can be interpreted as follows. In the high-noise regime when $\sigma \to \infty$, the output $\mathbf{Y}$ carries no information on $\mathbf{X}$ (i.e., $\mathbf{Y} \approx \mathbf{N}$) and hence, only the prior distribution on $\mathbf{X}$ can be used for the estimation of the permutation of $\mathbf{X}$. Thus, an optimal decoder would output the permutation that has the highest probability, which would incur the probability of error in (4.34). As we did in Section 4.4 for the low-noise regime, we consider the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ also for characterizing the high-noise asymptotics. As highlighted in Section 4.3 and Section 4.4, this decoder benefits from several appealing properties (e.g., low-complexity, optimal in certain scenarios). Using $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is also reasonable when the prior data distribution is not known. With this, (4.34) reduces to

$$\lim_{\sigma \to \infty} P_{e,\text{lin}}(\sigma) = 1 - \frac{1}{n!} = P_e(\infty). \tag{4.35}$$

The interpretation is that, since $\mathbf{Y} \approx \mathbf{N}$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, then any permutation of $\mathbf{N}$ is equally likely, i.e., it occurs with a probability of $1/n!$. Thus, the correct estimation of the permutation of $\mathbf{X}$ only happens when the permutation of $\mathbf{N}$ is equal to the permutation of $\mathbf{X}$, which leads to (4.35). In the remaining of this section, we study the rate of convergence of P_e in the high-noise regime (i.e., $\sigma \to \infty$).

4.5.1 Arbitrary Data Distribution

We here focus on the asymptotic behavior of the probability of error in Theorem 4.3.2 in the high-noise regime, without any assumption on the distribution of $\mathbf{X}$. In such a setting, as highlighted in Remark 4.4.2, the isotropic noise assumption (i.e., $K_{\mathbf{N}} = I_n$) is without loss of generality. The following theorem, proved in Appendix B.4, sharpens the limit in (4.35) by finding the rate of convergence in the high-noise regime.

Theorem 4.5.1. *Let $\mathbf{X} \in \mathbb{R}^n$ be a continuous random vector and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume*

that $\mathbb{E}[|X_i|] < \infty$ *for all* $i \in [1:n]$*, and that the linear decoder* $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ *is used. Then,*

$$\lim_{\sigma\to\infty} \frac{P_{e,\text{lin}}(\infty) - P_{e,\text{lin}}(\sigma)}{\frac{1}{\sigma}} = \frac{1}{\sqrt{2\pi}} \sum_{\tau\in\mathcal{P}} \sum_{i=1}^{n-1} \alpha_i P_{\mathbf{X}}(\mathcal{H}_\tau) \mathbb{E}\left[(\mathbf{W}_\tau)_i\right], \tag{4.36}$$

with $\mathbf{W}_\tau$ *being the* τ*-spacing of* $\mathbf{X}$ *in* (4.7) *and*

$$\alpha_i = \frac{\text{Vol}\left(\mathcal{E}(\mathbf{0}_{n-1}, i) \cap \mathcal{H}_{(1,2,\ldots,n-1)}\right)}{\text{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1)\right)}, \tag{4.37}$$

where $\mathcal{H}_{(1,2,\ldots,n-1)}$ *is defined in* (2.1)*,* $\mathcal{B}(\mathbf{0}_{n-1}, 1)$ *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.*

Remark 4.5.2. The constants α_i's in Theorem 4.5.1 can be expressed by a probability of a specific event on $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_{n-1}, I_{n-1})$. In particular (see also Appendix B.4),

$$\alpha_i = \Pr\left(Z_1 \leq \cdots \leq Z_{i-1} \leq \frac{1}{\sqrt{2}} Z_i \leq Z_{i+1} \leq \cdots \leq Z_{n-1} \right). \tag{4.38}$$

Finding a closed-form expression for the α_i's in (4.37), or in (4.38), does not appear to be an easy task. In the next proposition, we provide upper and lower bounds on the α_i's, which lead to expressions that are amenable to computations.

Proposition 4.5.3. *Let the assumptions in Theorem 4.5.1 hold. In the high-noise regime, the convergence rate of* $P_{e,\text{lin}}(\sigma)$ *in* (4.36) *is bounded as*

$$\frac{\mathbb{E}\left[R_n\right]}{\sqrt{\pi}(n-1)! 2^{\frac{n}{2}}} \leq \lim_{\sigma\to\infty} \frac{P_{e,\text{lin}}(\infty) - P_{e,\text{lin}}(\sigma)}{\frac{1}{\sigma}} \leq \frac{\mathbb{E}\left[R_n\right]}{\sqrt{2\pi}(n-1)!}, \tag{4.39}$$

where $R_n = X_{n:n} - X_{1:n}$ *is the range of* $\mathbf{X}$*.*

Proof. We start by observing that

$$\mathcal{B}\left(\mathbf{0}_{n-1}, 2^{-\frac{1}{2}}\right) \overset{\text{(i)}}{\subset} \mathcal{E}(\mathbf{0}_{n-1}, i) \overset{\text{(ii)}}{\subset} \mathcal{B}\left(\mathbf{0}_{n-1}, 1\right), \tag{4.40}$$

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

maximum radius equal to 1. Thus, from (4.40) we obtain

$$\alpha_i \leq \frac{\text{Vol}\left(\mathcal{B}\left(\mathbf{0}_{n-1},1\right)\cap\mathcal{H}_{(1,2,\ldots,n-1)}\right)}{\text{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1},1)\right)} = \frac{1}{(n-1)!}, \tag{4.41}$$

where the last equality follows since $\mathcal{H}_{(1,2,\ldots,n-1)}$ is a cone that occupies a $\frac{1}{(n-1)!}$ portion of the space and hence, $\text{Vol}\left(\mathcal{B}\left(\mathbf{0}_{n-1},1\right)\cap\mathcal{H}_{(1,2,\ldots,n-1)}\right) = \frac{1}{(n-1)!}\text{Vol}\left(\mathcal{B}\left(\mathbf{0}_{n-1},1\right)\right)$.

Similarly, from (4.40) we obtain

$$\begin{aligned}\alpha_i &\geq \frac{\text{Vol}\left(\mathcal{B}\left(\mathbf{0}_{n-1},2^{-\frac{1}{2}}\right)\cap\mathcal{H}_{(1,2,\ldots,n-1)}\right)}{\text{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1},1)\right)}\\ &= \left|\det\left(2^{-\frac{1}{2}}I_{n-1}\right)\right|\frac{\text{Vol}\left(\mathcal{B}\left(\mathbf{0}_{n-1},1\right)\cap\mathcal{H}_{(1,2,\ldots,n-1)}\right)}{\text{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1},1)\right)}\\ &= \frac{1}{2^{\frac{n-1}{2}}(n-1)!},\end{aligned} \tag{4.42}$$

where in the first equality we have used the facts that: (i) $2^{\frac{1}{2}}I_{n-1}\mathcal{B}\left(\mathbf{0}_{n-1},2^{-\frac{1}{2}}\right) = \mathcal{B}\left(\mathbf{0}_{n-1},1\right)$, (ii) $2^{\frac{1}{2}}I_{n-1}\mathcal{H}_{(1,2,\ldots,n-1)} = \mathcal{H}_{(1,2,\ldots,n-1)}$, and (iii) $\text{Vol}(A\mathcal{S}) = |\text{det}(A)|\text{Vol}(\mathcal{S})$ for any invertible matrix A and any set $\mathcal{S}$. The proof of Proposition 4.5.3 is concluded by substituting 4.41 and (4.42) inside (4.36) and by using the fact that

$$\begin{aligned}\sum_{\tau\in\mathcal{P}}\sum_{i=1}^{n-1}P_{\mathbf{X}}(\mathcal{H}_\tau)\mathbb{E}\left[(\mathbf{W}_\tau)_i\right] &= \sum_{\tau\in\mathcal{P}}P_{\mathbf{X}}(\mathcal{H}_\tau)\mathbb{E}\left[X_{n:n}-X_{1:n}\mid\mathbf{X}\in\mathcal{H}_\tau\right]\\ &= \mathbb{E}\left[X_{n:n}-X_{1:n}\right]\\ &= \mathbb{E}[R_n],\end{aligned} \tag{4.43}$$

where $R_n = X_{n:n} - X_{1:n}$ denotes the range of $\mathbf{X}$ [53]. □

We next analyze Theorem 4.5.1 and Proposition 4.5.3 for a case for which the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is optimal, i.e., $P_e(\sigma) = P_{e,\text{lin}}(\sigma)$.

4.5.2 Exchangeable Data Distribution

We here consider a distribution of $\mathbf{X}$ that is exchangeable, which implies that the linear decoder $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ is optimal (see Lemma 4.3.5). As shown in Section 4.4 for the low-noise regime,

also in the high-noise regime the exchangeability assumption provides a simplification for the probability of error expression. In particular, under the exchangeability assumption, we have that $\mathbb{E}[|X_i|]$ is the same for all $i \in [1:n]$ and hence, it suffices that there exists an $i \in [1:n-1]$ for which $\mathbb{E}[|X_i|] < \infty$. Moreover, since $\mathbf{W}_\tau \overset{d}{=} \mathbf{W}$, the probability of error in Theorem 4.5.1 becomes

$$\begin{aligned}\lim_{\sigma\to\infty} \frac{P_{e,\mathrm{lin}}(\infty) - P_{e,\mathrm{lin}}(\sigma)}{\frac{1}{\sigma}} &= \frac{1}{\sqrt{2\pi}} \sum_{\tau\in\mathcal{P}} \sum_{i=1}^{n-1} \alpha_i P_{\mathbf{X}}(\mathcal{H}_\tau) \mathbb{E}\left[(\mathbf{W}_\tau)_i\right] \\ &= \frac{1}{\sqrt{2\pi}} \sum_{\tau\in\mathcal{P}} \frac{1}{n!} \sum_{i=1}^{n-1} \alpha_i \mathbb{E}\left[W_i\right] \\ &= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \alpha_i \mathbb{E}\left[W_i\right]. \end{aligned} \tag{4.44}$$

The following corollary formally evaluates Theorem 4.5.1 for the case when $\mathbf{X}$ is exchangeable.

Corollary 4.5.4. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume that there exists an $i \in [1:n]$ for which $\mathbb{E}[|X_i|] < \infty$, and that the linear decoder $\phi_{\mathrm{lin}}(\cdot; I_n, \mathbf{0}_n)$ is used. Then, we have that $P_e(\sigma) = P_{e,\mathrm{lin}}(\sigma)$ with*

$$\lim_{\sigma\to\infty} \frac{P_{e,\mathrm{lin}}(\infty) - P_{e,\mathrm{lin}}(\sigma)}{\frac{1}{\sigma}} = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \alpha_i \mathbb{E}\left[W_i\right], \tag{4.45}$$

where $W_i = X_{i+1:n} - X_{i:n},\ i \in [1:n-1]$ and $\alpha_i, i \in [1:n]$ is defined in (4.37)*.*

Remark 4.5.5. Note that, as also highlighted at the beginning of Section 4.5, in the high-noise regime the dependence of the probability of error on the prior distribution of $\mathbf{X}$ is rather strong. This is indeed confirmed by Corollary 4.5.4, where it is shown that in the high-noise regime, the rate of the error probability depends on all the $n-1$ spacings.

We conclude this section by providing some evaluations of the expected spacing $\mathbb{E}[W_i]$ in Corollary 4.5.4 and of the range R_n in Proposition 4.5.3 for a few common distributions with i.i.d. data (see Appendix B.8.1 for the detailed computations). In particular, all these examples show that the term $\frac{1}{(n-1)!}$ dominates in the expression of the rate of the probability of error (see Proposition 4.5.3) for several distributions of interest.

Example 4.5.6. Consider $X \sim \text{Unif}(a, b),\ 0 \leq a < b < \infty$. Then,

$$\mathbb{E}[W_i] = \frac{b-a}{n+1} \text{ and } \mathbb{E}[R_n] = (b-a)\frac{(n-1)}{n+1}.$$

Example 4.5.7. Consider $X \sim \text{Exp}(\lambda),\ \lambda > 0$. Then,[3]

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

Example 4.5.8. Let X be γ^2-sub-Gaussian[4]. Then [62],

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

4.6 High-dimensional Regime

We here study the asymptotic behavior of the probability of error in the high-dimensional regime (i.e., $n \to \infty$). Different from the low-noise and high-noise asymptotics analyzed in Section 4.4 and Section 4.5, we here need some assumptions on the distribution of the input data vector $\mathbf{X}$. This is because n is a characteristic of $\mathbf{X}$, and hence, assuming an arbitrary distribution on $\mathbf{X}$ would make the problem intractable. Because of this, we here focus on the case when the components of $\mathbf{X}$ are i.i.d. random variables. We start by deriving the following theorem, proved in Appendix B.5, which presents lower and upper bounds on the probability of correctness.

Theorem 4.6.1. *Let $\mathbf{X}$ be an n-dimensional vector of i.i.d. random variables, $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, and let $\mathcal{X}_n$ be the chi random variable with n degrees of freedom. Then, $P_c(\sigma) = P_{c,\text{lin}}(\sigma)$ with*

$$\frac{1}{n!} + \frac{n!-1}{n!}\mathbb{E}\left[F_{\mathcal{X}_n}\left(\frac{\min_i\{W_i\}}{\sqrt{2}\sigma}\right)\right] \leq P_{c,\text{lin}}(\sigma), \tag{4.46}$$

and

$$P_{c,\text{lin}}(\sigma) \leq \prod_{i=1}^{n-1}\Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right), \tag{4.47}$$

[3]The quantity $\sum_{k=1}^{n-1}\frac{1}{k}$ is known as the harmonic number.

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

where $\mathbf{W}$ *is the spacing of* $\mathbf{X}$ *in Definition 4.2.1, and* $F_{\mathcal{X}_n}(\cdot)$ *and* $\Phi(\cdot)$ *denote the cumulative density function (cdf) of* $\mathcal{X}_n$ *and of the standard Gaussian random variable, respectively.*

Remark 4.6.2. If the support of $\mathbf{X}$ is bounded, the expected value of the spacing decreases as the dimension n increases, i.e., $\mathbb{E}[W_i] \to 0$ as $n \to \infty$. For an example of $\mathbf{X}$ with unbounded support see [63] where it was shown that the spacing of standard normal random variables converges to 0 as $n \to \infty$. This implies that, in the high-dimensional regime with a fixed $\sigma > 0$, we have $\Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right) \to \frac{1}{2}$, and thus $\lim_{n\to\infty} P_c(\sigma) = 0$ with an exponential rate.

The bounds on $P_c(\sigma)$ in Theorem 4.6.1 depend on the distribution of $\mathbf{X}$ through the spacing variables $W_i, i \in [1 : n-1]$. For instance, if $X \sim \text{Unif}(a, b),\ 0 \leq a < b < \infty$, then $\mathbb{E}[W_i] = (b-a)/(n+1)$ for all $i \in [1 : n-1]$ and hence, the upper bound in Theorem 4.6.1 reduces to

$$\begin{aligned} P_c(\sigma) = P_{c,\text{lin}}(\sigma) &\leq \prod_{i=1}^{n-1} \Phi\left(\frac{b-a}{\sqrt{2}\sigma(n+1)}\right) \\ &= \left[\Phi\left(\frac{b-a}{\sqrt{2}\sigma(n+1)}\right)\right]^{n-1}. \end{aligned} \tag{4.48}$$

Although the upper bound in (4.47) depends on the distribution of $\mathbf{X}$, we now compute a universal upper bound on $P_c(\sigma)$, which holds for a large class of distributions, namely the sub-Gaussian. The proof of Corollary 4.6.3 can be found in Appendix B.6.

Corollary 4.6.3. *Let* X *be* γ^2*-sub-Gaussian, and let the assumptions of Theorem 4.6.1 hold. Then,*

$$\begin{aligned} P_c(\sigma) = P_{c,\text{lin}}(\sigma) &\leq \left(\frac{1}{2} + \frac{\sqrt{2\gamma^2 \log(n)}}{\sigma\sqrt{\pi}(n-1)}\right)^{n-1} \\ &\leq 2^{-n+1} \mathrm{e}^{\frac{2\sqrt{2}\gamma}{\sqrt{\pi}\sigma}\sqrt{\log(n)}}. \end{aligned} \tag{4.49}$$

Theorem 4.6.1 allows us to draw some general conclusions, such as that $\lim_{n\to\infty} P_c(\sigma) = 0$ with an exponential rate. However, the bounds in (4.46) and (4.47) can be tightened if specific distributions are considered. We next show this fact by considering the case when $\mathbf{X}$ is an n-dimensional vector of i.i.d. Gaussian random variables. In particular, the next proposition, proved in Appendix B.7, shows that for this case the convergence rate of $P_c(\sigma)$ has $\frac{1}{n!}$ as dominant factor.

Proposition 4.6.4. *Assume that* $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. *Then, the probability of correctness* $P_c(\sigma) = P_{c,\mathrm{lin}}(\sigma)$ *can be upper and lower bounded as*

$$\frac{1}{n!} \leq P_{c,\mathrm{lin}}(\sigma) \leq \frac{1}{n!} \frac{\|A\|^{2n}}{\sigma^n}, \tag{4.50}$$

where

$$\|A\| = \left(\frac{(\sigma^4 + 4)^{\frac{1}{2}}}{2} + \frac{\sigma^2}{2} + 1 \right)^{\frac{1}{2}}. \tag{4.51}$$

Consequently,

$$\lim_{n \to \infty} \frac{\log P_{c,\mathrm{lin}}}{\log(\frac{1}{n!})} = 1. \tag{4.52}$$

The results in this section have shown a rate of convergence of $P_c(\sigma)$ that is at least exponential in the data dimension n. We now conclude this section with a remark that highlights that the result in Proposition 4.6.4 holds beyond the Gaussian assumption on the input data distribution.

Remark 4.6.5. The result in Proposition 4.6.4 holds whenever the pair $(\mathbf{X}, \mathbf{N})$ follows a spherically symmetric distribution. When $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, the assumption of a Gaussian i.i.d. input distribution satisfies this requirement.

4.7 Discussion and conclusion

In this chapter, we thoroughly investigated the error probability incurred by the data permutation recovery problem in a Gaussian noise setting when a linear decoder is used for the estimation task. In particular, in Section 4.3 we characterized the error probability, and then we analyzed its asymptotic behavior in the low-noise (Section 4.4), high-noise (Section 4.5), and high-dimensional (Section 4.6) regimes. Our results showcase that the permutation recovery problem is noise-dominated, i.e., if the noise standard deviation and the data dimension are not small enough, we are able to recover the data permutation with only approximately $\frac{1}{n!}$ probability.

We conclude this section with a few interesting future research directions:

1. *Approximate Permutation Recovery*: In some applications, one might be interested in recovering the data permutation up to a given distortion measured by a ranking distance function [64] (and not *exactly* as we have considered in this paper). In Chapter 6 and in [5], we have started this line of research and we have derived sufficient conditions that ensure a sub-linear (in the noise standard deviation) error behavior in the low-noise regime (and not linear as shown in Theorem 4.4.1).

2. *Partial Permutation Recovery*: Another research direction consists of investigating the problem of recovering the permutation of only part of a data vector, instead of the entire data vector. For instance, one might be interested in recovering the permutation of $\mathbf{X}_{\mathcal{I}}$, where $\mathbf{X}_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}$ is a sub-vector of $\mathbf{X} \in \mathbb{R}^n$ indexed by a set $\mathcal{I} \subset [1:n]$. Reducing the 'target' data dimension might have a significant impact on the difficulty of the problem, as can also be observed from Theorem 4.6.1 and Proposition 4.6.4. In particular, the interesting scenario to study would be when the data has memory since the i.i.d. setting is a trivial sub-problem of the *exact* permutation recovery studied in this paper.

3. *Characterizing the Best Linear Decoder*: In Chapter 3, we established conditions on the noise and data distributions under which a linear decoder is optimal. However, given the appealing properties of the linear decoder, it would be interesting to apply it beyond such conditions. Thus, a natural question arises: What is the best (i.e., the one that incurs the minimum probability of error) linear decoder? In other words, what is a solution to the following optimization problem?

$$\underset{A\in\mathbb{R}^{n\times n},\, \mathbf{b}\in\mathbb{R}^n}{\text{minimize}} \quad P_{e,\text{lin}}(A,\mathbf{b}), \tag{4.53}$$

   where $P_{e,\text{lin}}(A,\mathbf{b}) = \sum_{\tau\in\mathcal{P}} \Pr(\mathbf{X} \notin \mathcal{H}_\tau, \mathbf{Y} \in A\mathcal{H}_\tau + \mathbf{b})$. Even for the case of Gaussian $\mathbf{X}$ and $\mathbf{N}$, the MMSE estimator is not always the optimal linear decoder for every covariance matrix, as shown by examples in [20].

4. *Memory Effect*: How does memory (i.e., correlation among the entries of $\mathbf{X}$) affect the permutation recovery problem? We conjecture that, while a 'negative' memory can indeed be helpful, a 'positive' memory makes the problem harder to solve. To support our intuition, consider Fig. 4.3, where we draw the ellipses $\mathcal{E}_{K_{\mathbf{X}}} = \left\{\mathbf{x} : \mathbf{x}^T K_{\mathbf{X}}^{-1}\mathbf{x} = 1\right\}$ parameterized by the 2×2 covariance matrix $K_{\mathbf{X}}$. Each ellipse represents a contour line

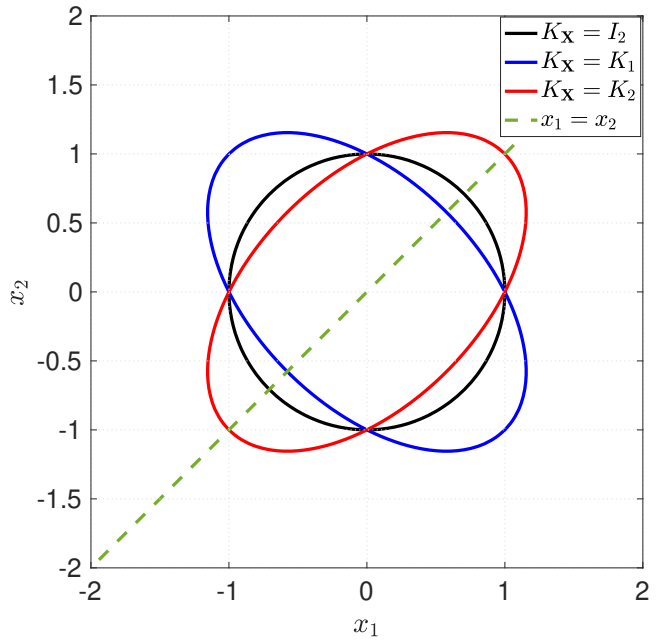


Figure 4.3: Ellipses $\mathcal{E}_{K_{\mathbf{X}}}$ corresponding to $K_{\mathbf{X}} \in \{I_2, K_1, K_2\}$, where $K_1 = \left[\begin{smallmatrix} 1 & -0.5 \\ -0.5 & 1 \end{smallmatrix}\right]$ and $K_2 = \left[\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right]$.

of the probability density of a bivariate Gaussian random vector with zero-mean and covariance matrix $K_{\mathbf{X}}$. In the permutation recovery problem with $n = 2$, it is not difficult to see that the permutation of a data vector close to the 'boundary' (i.e., $x_1 = x_2$) is more difficult to be correctly decoded than the one of a data vector that is farther away from the boundary. This is because a data vector close to the boundary can be shifted to the opposite region by adding noise (e.g., a standard Gaussian noise) with a higher probability than a data vector that is farther away from the boundary. From this observation, we infer that a 'positive' memory in the data vector makes the problem harder to solve as also empirically shown in Fig. 4.1, while a 'negative' memory can indeed be helpful. As highlighted throughout the paper, the spacing (see Section 4.2) plays a critical role in the permutation recovery problem. Thus, investigating the connection between spacing and memory could be useful to better understand the problem on the memory effect. Regarding this, it is worth noting that for exchangeable random variables, it was shown in [65] that if the random variables have a positive memory, then the spacing vector becomes stochastically smaller [66]. This result supports our conjecture on the memory effect.

Chapter 5

Permutation Recovery under Privacy Considerations

5.1 Introduction

In this chapter, we study the *private ranking recovery* problem, which consists of recovering the ranking/permutation of an input data vector from a noisy version of it. The importance and timeliness of this problem stems from two major considerations. First, many modern computing systems are often more interested in recovering the permutation, i.e., the relative ranking of data points, rather than the values of the data itself. Second, because of privacy considerations, users might decide to privatize their data (e.g., by adding some noise) before sharing it with an external party. These facts give rise to the following practically relevant question: *Which perturbation mechanisms allow for data privatization, while still allowing to correctly recover the permutation of the input data vector with high probability?*

5.1.1 Related work

Problems with a similar flavor to the private ranking recovery problem have been analyzed in literature. The rank aggregation problem, the goal of which is to find a representative ranking for multiple data rankings, was studied under differential privacy constraints by [37] and [38], and under local differential privacy by [39] and [40]. *Differential privacy* (DP), which is a statistical guarantee introduced by [67] for indistinguishability whether any data element exists

or not in a dataset, is one of the most common adopted privacy metrics. Several notions of DP have been introduced and analyzed that range from the basic ϵ-DP metric (which can be guaranteed by using the Laplace randomized mechanism) [16], to more relaxed versions of it, such as the (ϵ, δ)-DP [68], the ϵ-mutual information DP and the ϵ-Kullback-Leibler (KL) DP [18], and the (α, ϵ)-Rényi DP (RDP) [17]. In particular, the (α, ϵ)-RDP encompasses: (i) the ϵ-DP if $\alpha \to \infty$, and (ii) the ϵ-KL DP if $\alpha \to 1$. Moreover, some important properties of the ϵ-DP, e.g., the composition theorem, remain applicable in the RDP framework.

When data is confidential, it needs to be privatized before being shared with an external party (which will perform some operations on it) a natural question arises: *For a fixed target performance guarantee (a.k.a. utility) required on the data, what is a randomized mechanism that achieves the maximum level of privacy?* To answer this question, one needs to understand the *trade-off* between privacy and utility. Such a trade-off has been studied in the literature in several settings, where different utility measures have been used. For instance, [69] compared several randomized mechanisms (from a statistical point of view) by using the Kolmogorov–Smirnov and the L_2 distances among distributions and densities. [70] showed a trade-off between the convergence of a federated learning algorithm (utility) and the level of privacy (measured in terms of DP) that can be guaranteed, hence suggesting the amount of artificial noise that should be used in this context. [71] studied the privacy-utility trade-off for a hyperparameterized algorithm using multi-objective optimization and Pareto front. For a single real-valued query, [72] identified the staircase distribution (i.e., a geometric mixture of uniform random variables) as a distribution that minimizes the L_1 loss (utility) under a fixed given level of ϵ-DP. The staircase distribution has also been shown to be the optimal ϵ-DP mechanism under other utility constraints [73]. More recently, [74] studied trade-offs between $(0, \delta)$-DP and the L_p loss function for a single real-valued query function.

5.1.2 Contributions

First, we formulate the private ranking recovery problem within a DP framework. In particular, we adopt the (α, ϵ)-RDP as a privacy metric; our choice mainly stems from the fact that the (α, ϵ)-RDP encompasses other widely employed DP metrics such as the ϵ-DP [16] if $\alpha \to \infty$, and the ϵ-KL DP [18] if $\alpha \to 1$. Moreover, as pointed out in [17, Proposition 3], (α, ϵ)-RDP can be converted to (ϵ, δ)-DP.

Second, we show that under mild assumptions on the input data vector (i.e., the input data

Table 5.1: Trade-off between privacy and utility in the low-noise regime with i.i.d. noise components. Privacy is measured by (α, ϵ)-RDP for the Gaussian and Laplace mechanisms and by ϵ-DP for the generalized normal mechanism. The utility is quantified by P_e.

$\mathcal{K}(\sigma)$	**Trade-off**
$\mathcal{N}(0,1)$	$P_e \propto \left(\frac{\alpha}{\epsilon}\right)^{1/2}$
$\text{Lap}\left(0, \frac{1}{\sqrt{2}}\right)$	$P_e \propto \frac{1}{\epsilon}$
$\mathcal{GN}\left(0, \sqrt{\frac{\Gamma(p^{-1})}{\Gamma(3p^{-1})}}, p\right)$	$P_e \propto \left(\frac{1}{\epsilon}\right)^{1/p}$

distribution is exchangeable) and on the randomized mechanism (i.e., it has an ℓ_p-spherical distribution), declaring the permutation of the observed noisy vector is an optimal decision rule for recovering the permutation of the input data vector. Because of this, and using the terminology introduced in [2, 3], we refer to such a decision rule as *linear decoder*. This has complexity $O(n \log n)$, which is a significant reduction with respect to the $O(n!)$ complexity of a naive brute-force implementation of the optimal decoder.

Third, we characterize the error probability of the linear decoder, by deriving the Taylor series of it. This result suggests that the private ranking recovery problem is noise dominated, i.e., the error probability is large even for small values of the noise variance. Further, we derive the first-order approximation of the error probability with respect to the noise standard deviation, and we verify through numerical simulations that this approximation is indeed accurate. In particular, our first-order approximation expression decouples the effects of the input data distribution and noise distribution on the error probability. We also derive the exact expression for the linear slope of the error probability for the case of i.i.d. input data vector entries.

Finally, we derive the trade-off between privacy (measured by ϵ-DP and (α, ϵ)-RDP) and utility (measured by the error probability P_e) in the low-noise regime. We consider widely used noise addition mechanisms, i.e., the Laplace, the Gaussian, and the generalized normal. As indicated in Table 5.1, these mechanisms have different relationships[1] between ϵ and P_e. The trade-offs for $\mathcal{N}(0,1)$ and $\text{Lap}\left(0, \frac{1}{\sqrt{2}}\right)$ are obtained based on (α, ϵ)-RDP, and for the generalized normal mechanism with $p \leq 1$, ϵ-DP is considered. We observe that the generalized

[1]The probability of error P_e is proportional up to the first-order term $\left(\frac{1}{\epsilon}\right)^{1/p}$.

normal mechanism with $p \leq 1$ offers the best trade-off.

5.1.3 Notation

Upon the notations introduced in Chapter 2.1. throughout this chapter, we use the following notation.

$\mathbb{1}_{\mathcal{S}}$ is the indicator function over the set $\mathcal{S}$. For any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, the Hamming distance is defined as $d_H(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq y_i\}}$; $\stackrel{d}{=}$ denotes equality in distribution; $\mathcal{N}(\boldsymbol{\mu}_n, K)$ is the n-dimensional Gaussian distribution with mean $\boldsymbol{\mu}_n$ and covariance matrix K; Lap(μ, b) is the Laplace distribution with mean μ and scale b; $\mathcal{GN}(\mu, a, p)$ is the generalized normal distribution [75, 76] with mean μ, scale a, and shape p. We let $\mathcal{P}$ be the set of all permutations of an n-dimensional vector. For $\tau \in \mathcal{P}$, recall that

$$\mathcal{H}_\tau = \{\mathbf{x} \in \mathbb{R}^n : x_{\tau_1} \leq x_{\tau_2} \leq \cdots \leq x_{\tau_n}\}, \tag{5.1}$$

with $x_{\tau_i}, i \in [1:n]$ being the τ_i-th element of $\mathbf{x}$, and $\tau_i, i \in [1:n]$ being the i-th element of τ. For example, in the 3-dimensional space there exist $|\mathcal{P}| = 6$ permutations, and we have

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

where $X_i, i \in [1:3]$ is the i-th element of $\mathbf{X}$.

5.2 Problem Formulation

We consider the private ranking recovery problem, as shown in Figure 5.1. In this setting, because of privacy considerations, a randomized mechanism $\mathcal{K}(\cdot)$ is applied on the confidential n-dimensional data vector $\mathbf{X} \in \mathbb{R}^n$, before this data is collected by an external party (e.g., recommender system). In other words, $\mathcal{K}(\cdot)$ is applied so as to hide the values of $\mathbf{X}$ from the collector (i.e., privatize $\mathbf{X}$). The goal of the data collector is then to retrieve the permutation $\pi_{\mathbf{X}}$ according to which $\mathbf{X}$ is sorted, i.e., to output the estimate $\hat{\pi}_{\mathbf{X}}$.

In the framework described above, a natural trade-off arises between the performance of the

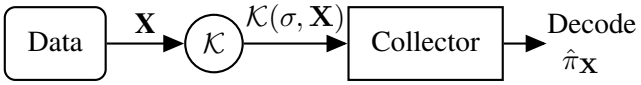


Figure 5.1: Graphical representation of the considered private ranking recovery framework.

estimation task, referred to as utility function in the remaining of this paper, and the privacy level that can be guaranteed. In particular, such a trade-off is dictated by the distribution of $\mathbf{X}$, and $\mathcal{K}(\cdot)$. In this work, we are interested in characterizing such a trade-off for randomized mechanisms that consist of noise addition on the data vector $\mathbf{X}$, namely

$$\mathcal{K}(\sigma, \mathbf{X}) \triangleq \mathbf{X} + \sigma\mathbf{N}, \tag{5.2}$$

where $\mathbf{N} \in \mathbb{R}^n$ is the n-dimensional noise random vector and $\sigma \geq 0$ is a parameter controlling the power of the noise.

Utility Function. As utility function, we consider the *probability of error* incurred in the estimation of $\pi_{\mathbf{X}}$. With reference to Figure 5.1, we let $\phi(\cdot) : \mathbb{R}^n \to \mathcal{P}$ denote the decoder that the data collector uses to output $\hat{\pi}_{\mathbf{X}}$. Then, the probability of error of the estimation task depends both on $\phi(\cdot)$ and $\mathcal{K}(\cdot)$, that is

$$P_e(\phi, \mathcal{K}) = \Pr\left(\phi(\mathcal{K}(\sigma, \mathbf{X})) \neq \pi_{\mathbf{X}}\right). \tag{5.3}$$

Privacy Metric. Given $\mathcal{K}(\sigma, \mathbf{X})$ in (5.2), it is important to quantify the privacy level guaranteed by this mechanism. Towards this end, we leverage the ϵ-DP [16] in Definition 5.2.1 and the (α, ϵ)-RDP [17] in Definition 5.2.2.

Definition 5.2.1. Let $\mathcal{X}$ be the set of possible n-dimensional real-valued data vectors. Let $(\mathbf{X}, \tilde{\mathbf{X}}) \in \mathcal{X}^2$ be a pair of adjacent data vectors, which differ in at most one element, i.e., $d_H(\mathbf{X}, \tilde{\mathbf{X}}) \leq 1$. Then, the randomized mechanism $\mathcal{K}(\cdot)$ gives ϵ-DP if, for any set $\mathcal{S}$, we have that

$$\Pr(\mathcal{K}(\sigma, \mathbf{X}) \in \mathcal{S}) \leq e^{\epsilon} \Pr(\mathcal{K}(\sigma, \tilde{\mathbf{X}}) \in \mathcal{S}). \tag{5.4}$$

Definition 5.2.2. Let $\mathcal{X}$ be the set of possible n-dimensional real-valued data vectors. Let $(\mathbf{X}, \tilde{\mathbf{X}}) \in \mathcal{X}^2$ be a pair of adjacent data vectors, which differ in at most one element, i.e.,

$d_H(\mathbf{X}, \tilde{\mathbf{X}}) \leq 1$. Then, for $\alpha \geq 1$, the randomized mechanism $\mathcal{K}(\cdot)$ gives (α, ϵ)-RDP if

$$\mathsf{RDP}_\alpha(\mathcal{K}) \leq \epsilon, \tag{5.5a}$$

where

$$\mathsf{RDP}_\alpha(\mathcal{K}) = \sup_{(\mathbf{X},\tilde{\mathbf{X}})\in\mathcal{X}^2 : d_H(\mathbf{X},\tilde{\mathbf{X}})\leq 1} D_\alpha(\mathcal{K}(\sigma, \mathbf{X}) \| \mathcal{K}(\sigma, \tilde{\mathbf{X}})), \tag{5.5b}$$

and, for $\mathbf{X}$ and $\mathbf{Y}$ with equal support,

$$D_\alpha(\mathbf{X}\|\mathbf{Y}) = \frac{1}{\alpha - 1} \log \mathbb{E}\left[\left(\frac{f_{\mathbf{X}}(\mathbf{Y})}{f_{\mathbf{Y}}(\mathbf{Y})}\right)^\alpha\right], \tag{5.5c}$$

with $f_{\mathbf{X}}(\cdot)$ and $f_{\mathbf{Y}}(\cdot)$ being the probability density functions (PDFs) of $\mathbf{X}$ and $\mathbf{Y}$, respectively. $D_\alpha(\cdot\|\cdot)$ is the Rényi divergence of order α.

Several rationales are behind our choice of using the (α, ϵ)-RDP as a privacy measure. First, the (α, ϵ)-RDP encompasses other widely employed DP metrics, e.g., the ϵ-DP [16] if $\alpha \to \infty$, and the ϵ-KL DP [18] if $\alpha \to 1$. The (α, ϵ)-RDP also bypasses some limitations of the ϵ-DP (e.g., a Gaussian noise adding mechanism is not ϵ-DP), while still retaining similar appealing properties (e.g., composition properties [17]) as those of the ϵ-DP.

Our goal in this paper is to characterize the privacy-utility trade-off when the randomized mechanism in (5.2) is used. In other words, we seek to determine $P_e(\phi, \mathcal{K})$ in (5.3), subject to the constraint that $\mathsf{RDP}_\alpha(\mathcal{K})$ in (5.5) is set to be equal to ϵ (for ϵ-DP we set $\alpha = \infty$). In particular, we will focus on scenarios where $\mathbf{X}$ is exchangeable and $\mathbf{N} \in \mathcal{S}_{n,p}$, as defined below.

Definition 5.2.3. A sequence of random variables $X_1, \ldots, X_n$ is said to be exchangeable if, for any permutation $\pi = (\pi_1, \ldots, \pi_n)$ of $[1 : n]$, we have

$$(X_1, \ldots, X_n) \overset{d}{=} (X_{\pi_1}, \ldots, X_{\pi_n}).$$

Definition 5.2.4. A function f is ℓ_p-spherically non-increasing if it can be written as

$$f(\mathbf{x}) = g(\|\mathbf{x}\|_p), \tag{5.6}$$

where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-increasing function. We denote by $\mathcal{S}_{n,p}$ the set of n-dimensional

distributions which have an ℓ_p-spherically non-increasing density function.

Our assumption on $\mathbf{X}$ being exchangeable includes data that does not need to be necessarily i.i.d., but can be correlated. For instance, any convex combination of i.i.d. random variables, and any spherically contoured distribution are exchangeable.[2] We also highlight that $\mathbf{N} \in \mathcal{S}_{n,p}$ implies that $\mathbf{N}$ is exchangeable; this follows since the ℓ_p-norm is permutation invariant. Finally, we conclude this section with a few examples (see Appendix C.1 for the details), which show that distributions on $\mathbf{N}$ widely used in the DP literature are in $\mathcal{S}_{n,p}$. Thus, the assumption that $\mathbf{N} \in \mathcal{S}_{n,p}$ can also be considered as mild.

Example 5.2.5. The following distributions belong to $\mathcal{S}_{n,p}$:

- $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$: in this case, $p = 2$;
- $\mathbf{N}$ consists of i.i.d. Lap$(0, b)$: in this case, $p = 1$;
- $\mathbf{N}$ consists of i.i.d. $\mathcal{GN}(0, a, p)$;
- $\mathbf{N}$ has a staircase distribution [72]: in this case, $p = 1$;
- $\mathbf{N} \sim$ Unif $(\mathcal{B}_p(\mathbf{0}_n, r))$ with $r > 0$, where $\mathcal{B}_p(\mathbf{0}_n, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_p < r\}$ is the ℓ_p-ball centered at $\mathbf{0}_n$.

5.3 Accuracy of Ranking Recovery

In this section, we seek to derive an expression for the probability of error of estimating $\pi_{\mathbf{X}}$. In Section 5.3.1, we first revisit a low-complexity decoder, and show its optimality under the assumptions of Section 5.2. Then, in Section 5.3.2 we characterize $P_e(\phi, \mathcal{K})$ for this decoder. In Section 5.3.3, we derive an accurate first-order approximation of $P_e(\phi, \mathcal{K})$, which we will leverage to characterize the privacy-utility trade-offs. Finally, in Section 5.3.4, we evaluate the derived first-order approximation of $P_e(\phi, \mathcal{K})$ for the case when the data $\mathbf{X}$ is i.i.d. and n is large (i.e., the high-dimensional regime).

[2]This restriction can be thought of as a limitation of our results. However, to make progress on this problem in a Bayesian framework, making assumptions is eventually inevitable as otherwise, the problem becomes intractable, and one will not be able to say much about the limits of permutation recovery. Assuming an exchangeable data distribution is reasonable whenever the data has no natural order. A particular example is relational data such as social network users, ratings, and preference data [61]. The exchangeability assumption, in our opinion, strikes a good balance between how permutation recovery would behave in practice and the problem theoretical solvability.

5.3.1 Optimal Decoder with Low-Complexity

As illustrated in Section 5.2, the data collector uses a decoder $\phi(\cdot) : \mathbb{R}^n \to \mathcal{P}$ to output $\hat{\pi}_{\mathbf{X}}$. In what follows, we let $\phi_{\text{opt}}(\cdot)$ denote the *optimal* decoder, i.e., the decoder that recovers $\hat{\pi}_{\mathbf{X}}$ such that the probability of error defined in (5.3) is minimized. We also consider a (potentially sub-optimal) decoder to which we refer as *linear decoder* and formally define below.

Definition 5.3.1. Given the noisy data vector $\mathbf{y} \in \mathbb{R}^n$, the linear decoder is defined as

$$\phi_{\text{lin}}(\mathbf{y}) = \pi_{\mathbf{y}}, \tag{5.7}$$

where $\pi_{\mathbf{y}}$ denotes the permutation according to which $\mathbf{y}$ is sorted.

The decoder in (5.7) is a special case of a more general linear decoder $\pi_{A\mathbf{y}+\mathbf{b}}$, where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$; such a linear decoder can be optimal when the noise has memory [22, 21, 2]. In (5.7), we set $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$.

The linear decoder $\phi_{\text{lin}}(\cdot)$ has several advantages, among which its low-complexity: it simply consists of a sorting operation and hence, it has a complexity of $O(n \log n)$. This is a significant reduction with respect to the $O(n!)$ complexity of a naive brute-force implementation of the optimal decoder $\phi_{\text{opt}}(\cdot)$ based on the maximum a posteriori (MAP) decision rule [54]. Moreover, as we will show in Theorem 5.3.3, the linear decoder $\phi_{\text{lin}}(\cdot)$ is indeed optimal (i.e., $\phi_{\text{lin}}(\cdot) = \phi_{\text{opt}}(\cdot)$) under the assumptions stated in Section 5.2. In particular, to show this result we will leverage the following lemma (proof in Appendix C.2).

Lemma 5.3.2. *For any two n-dimensional vectors $\mathbf{x} \in \mathcal{H}_\eta$ and $\mathbf{y} \in \mathcal{H}_\tau$, and $p \geq 1$, we have that*

$$\tau \in \arg\min_{\omega\in\mathcal{P}} \|\mathbf{y} - P_{\eta\to\omega}\mathbf{x}\|_p, \tag{5.8}$$

where $P_{\eta\to\omega}$ is the permutation matrix that permutes $\mathbf{x} \in \mathcal{H}_\eta$ into $P_{\eta\to\omega}\mathbf{x} \in \mathcal{H}_\omega$.

Lemma 5.3.2 states that, when $p \geq 1$, the ℓ_p-norm of the difference between two given vectors is minimized when the two vectors are sorted according to the same permutation. Lemma 5.3.2 allows us to prove our first main result, which is given by the next theorem.

Theorem 5.3.3. *Let $\mathbf{X} \in \mathbb{R}^n$ be exchangeable, and assume that the randomized mechanism $\mathcal{K}(\sigma, \mathbf{X})$ adopts $\mathbf{N} \in \mathcal{S}_{n,p},\ p \geq 1$. Then, given any noisy data vector $\mathbf{y} \in \mathbb{R}^n$, we have that*

$$\phi_{\text{opt}}(\mathbf{y}) = \phi_{\text{lin}}(\mathbf{y}). \tag{5.9}$$

Proof. Since $\mathbf{X}$ is exchangeable, all hypotheses are equally-likely (i.e., $\Pr(\mathbf{X} \in \mathcal{H}_\tau) = \frac{1}{n!}$, $\forall \tau \in \mathcal{P}$), and the maximum likelihood decoder is optimal [54]. This can be shown as follows,

$$\begin{aligned}\phi_{\text{opt}}(\mathbf{y}) &= \arg\max_{\tau\in\mathcal{P}} \Pr(\mathbf{X} \in \mathcal{H}_\tau \mid \mathcal{K}(\sigma, \mathbf{X}) = \mathbf{y}) \\ &= \arg\max_{\tau\in\mathcal{P}} \frac{\Pr(\mathbf{X} \in \mathcal{H}_\tau)}{f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y})} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau) \\ &= \arg\max_{\tau\in\mathcal{P}} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau), \end{aligned} \tag{5.10}$$

where $f_{\mathcal{K}(\sigma,\mathbf{X})}$ is the PDF of $\mathcal{K}(\sigma, \mathbf{X})$. We note that the second equality follows by the Bayes' rule, and the last equality follows by the facts that $\Pr(\mathbf{X} \in \mathcal{H}_\tau)$ is a constant for all $\tau \in \mathcal{P}$ and that $f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y})$ is independent of τ. Therefore, given $\mathbf{y} \in \mathbb{R}^n$ for $\mathcal{K}(\sigma, \mathbf{X})$, an optimal decoder is given by

$$\phi_{\text{opt}}(\mathbf{y}) = \arg\max_{\tau\in\mathcal{P}} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau). \tag{5.11}$$

Since $\mathbf{X}$ and $\mathbf{N}$ are independent, the conditional density function in (5.11) can be written as

$$\begin{aligned} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau) &= \int f_{\mathbf{X}}(\mathbf{x} \mid \mathcal{H}_\tau) f_{\mathbf{N}}(\mathbf{y} - \mathbf{x})\, \mathrm{d}\mathbf{x} \\ &= n! \int \mathbb{1}_{\{\mathbf{x}\in\mathcal{H}_\tau\}} f_{\mathbf{X}}(\mathbf{x}) g(\|\mathbf{y} - \mathbf{x}\|_p)\, \mathrm{d}\mathbf{x}, \end{aligned} \tag{5.12}$$

where in the last equality we used Definition 5.2.4 with $g(\cdot)$ being a non-increasing function. Similarly, we have

$$\begin{aligned} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\eta) &= n! \int \mathbb{1}_{\{\mathbf{x}\in\mathcal{H}_\eta\}} f_{\mathbf{X}}(\mathbf{x}) g(\|\mathbf{y} - \mathbf{x}\|_p)\, \mathrm{d}\mathbf{x} \\ &= n! \int \mathbb{1}_{\{\mathbf{u}\in\mathcal{H}_\tau\}} f_{\mathbf{X}}(\mathbf{u}) g(\|\mathbf{y} - P_{\tau\to\eta}\mathbf{u}\|_p)\, \mathrm{d}\mathbf{u}, \end{aligned} \tag{5.13}$$

where (5.13) follows by substituting $\mathbf{x} = P_{\tau\to\eta}\mathbf{u}$.

Now, by taking the difference between (5.12) and (5.13), we obtain

$$\begin{aligned} &\frac{1}{n!}\left(f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau) - f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\eta)\right) \\ &= \int_{\mathbf{x}\in\mathcal{H}_\tau} f_{\mathbf{X}}(\mathbf{x}) \left(g(\|\mathbf{y} - \mathbf{x}\|_p) - g(\|\mathbf{y} - P_{\tau\to\eta}\mathbf{x}\|_p)\right) \mathrm{d}\mathbf{x}. \end{aligned} \tag{5.14}$$

Using Lemma 5.3.2, we have that if $\mathbf{y} \in \mathcal{H}_\tau$, then the integrand in (5.14) is always non-negative.

Hence, for any $\mathbf{y}$ sorted according to $\pi_{\mathbf{y}}$, we have

$$\phi_{\text{opt}}(\mathbf{y}) = \arg\max_{\tau \in \mathcal{P}} f_{\mathcal{K}(\sigma,\mathbf{X})}(\mathbf{y} \mid \mathbf{X} \in \mathcal{H}_\tau) = \pi_{\mathbf{y}} = \phi_{\text{lin}}(\mathbf{y}),$$

where the last equality follows from Definition 5.3.1. This concludes the proof of Theorem 5.3.3. □

Remark 5.3.4. We highlight that chapter 3 showed a similar result as in Theorem 5.3.3 for the case of Gaussian noise, under some specific conditions on the noise covariance matrix. Theorem 5.3.3 extends the result on the optimality of the linear decoder beyond Gaussian noise, i.e., whenever $\mathbf{N} \in \mathcal{S}_{n,p},\ p \geq 1$. In particular, to show this result we have leveraged a completely new proof which uses a generalized version of the rearrangement inequality needed in the proof of Lemma 5.3.2 (see Appendix C.2).

5.3.2 Error Analysis for $\phi_{\text{lin}}(\cdot)$

We here characterize the error probability of the low-complexity and optimal (as proved in Theorem 5.3.3 under some assumptions) decoder $\phi_{\text{lin}}(\cdot)$ in Definition 5.3.1. From (5.3), the error probability when $\phi_{\text{lin}}(\cdot)$ is used is

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \Pr(\phi_{\text{lin}}(\mathcal{K}(\sigma, \mathbf{X})) \neq \pi_{\mathbf{X}}).$$

Before deriving $P_e(\phi_{\text{lin}}, \mathcal{K})$, we use the matrix $T_\tau \in \mathbb{R}^{(n-1)\times n}$ defined in (4.4). Specifically, for all $\tau \in \mathcal{P}$, T_τ is defined as

$$(T_\tau)_{i,j} = \mathbb{1}\{j = \tau_{i+1}\} - \mathbb{1}\{j = \tau_i\}. \tag{5.15}$$

For instance, let $n = 4$ and $\tau = (4, 2, 1, 3)$; then,

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

Remark 5.3.5. For any exchangeable $\mathbf{X} \in \mathbb{R}^n$, we have that [60]

$$T_\tau \mathbf{X} \mid \mathbf{X} \in \mathcal{H}_\tau \stackrel{d}{=} \mathbf{W},\ \forall \tau \in \mathcal{P}, \tag{5.16a}$$

where $\mathbf{W} \in \mathbb{R}^{n-1}$ is known as the spacing vector [53] with

$$W_i \stackrel{d}{=} X_{i+1:n} - X_{i:n},\ i \in [1 : n-1], \tag{5.16b}$$

where $X_{i:n}$ is the i-th order statistics of $\mathbf{X}$.

The theorem below provides an expression for the error probability of the private ranking recovery problem when the linear decoder $\phi_{\text{lin}}(\cdot)$ in Definition 5.3.1 is used. In particular, this expression is derived by considering the Taylor series of the error probability at $\sigma = 0$.

Theorem 5.3.6. *Assume that* $\lim_{\sigma \to 0^+} |f^{(i)}_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{w})| < \infty$, *for all* $\mathcal{I} \subseteq [1 : n-1]$ *where* $f^{(i)}_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{w}) := \frac{\partial^i}{\partial \sigma^i} f_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{w})$. *Then, the Taylor series of* $P_e(\phi_{\text{lin}}, \mathcal{K})$ *is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \sum_{i=0}^{\infty} \frac{P_e^{(i)}}{i!} \sigma^i, \tag{5.17}$$

where

$$P_e^{(i)} = \sum_{k=1}^{\min\{i, n-1\}} (-1)^{k-1} \binom{i}{k} k! \alpha_k^{(i-k)}(0^+),$$

and

$$\alpha_k^{(i-k)}(\omega) = \sum_{\substack{\mathcal{I} \subseteq [1:n-1] \\ |\mathcal{I}| = k}} \int_{\mathbf{u} \in \mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}(-\mathbf{u}) f^{(i-k)}_{\mathbf{W}_\mathcal{I}}(\omega \mathbf{u}) \mathrm{d}\mathbf{u},$$

where $F_{\mathbf{V}_\mathcal{I}}(\cdot)$ *is the cumulative distribution function (CDF) of* $\mathbf{V}_\mathcal{I}$ *with* $V_i = N_{i+1} - N_i$ *for* $i \in [1 : n-1]$.

We defer the proof of Theorem 5.3.6 to Appendix C.3. Note that Theorem 5.3.6 (and also the following Corollary 5.3.7) generalizes Theorem 4.4.1 under two aspects: (i) from the first-order coefficient to an arbitrary order coefficient; and (ii) beyond Gaussian noise.

As an application of Theorem 5.3.6, we next present a corollary (proof in Appendix C.4),

which provides the second-order approximation of $P_e(\phi_{\text{lin}}, \mathcal{K})$ for $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$ and any exchangeable distribution of $\mathbf{X}$.

Corollary 5.3.7. *Let $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Assume that, for $i, j \in [1 : n-1]$, $|f'_{W_i}(w)| < \infty$, $\forall w$ and $f_{W_i,W_j}(u, v) < \infty$, $\forall (u, v)$. Then, a second order approximation of P_e in the low-noise regime is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}_{\mathbf{N}}) = c_1\sigma + c_2\sigma^2 + O(\sigma^3), \tag{5.18}$$

where [3]

$$c_1 = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}},$$
$$c_2 \approx \frac{1}{2}\sum_{i=1}^{n-1} f'_{W_i}(0^+) - 0.108998 \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) - \frac{1}{\pi} \sum_{\substack{(i,j)\in[1:n-1]^2 \\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+).$$

We note that the constants c_1 and c_2 in Corollary 5.3.7 depend on the distribution of $\mathbf{X}$. Next, as an example, we derive closed-form expressions for c_1 and c_2 for $X_i \sim \text{Unif}(0, 1)$ and $X_i \sim \text{Exp}(\lambda)$. The detailed proof of these examples can be found in Appendix C.5, where we also provide various simulation results that graphically showcase the accuracy of the result in Corollary 5.3.7.

Example 5.3.8. Let $X_i \sim \text{Unif}(0, 1)$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Then, the constants c_1 and c_2 in Corollary 5.3.7 are

$$c_1 = \frac{n(n-1)}{\sqrt{\pi}},$$
$$c_2 \approx -\frac{1}{2}n(n-1)^2 - 0.108998n(n-1)(n-2) - \frac{1}{2\pi}n(n-1)(n-2)(n-3).$$

Example 5.3.9. Let $X_i \sim \text{Exp}(\lambda)$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. Then, the constants c_1 and c_2 in

[3] The approximation of c_2 can be made exact by replacing 0.108998 with its exact value $\mathbb{E}[\max\{0, V_1\}\max\{0, V_2\}]$ where V_i's are defined in Thereom 5.3.6.

Corollary 5.3.7 are

$$c_1 = \frac{n(n-1)\lambda}{2\sqrt{\pi}},$$
$$c_2 \approx -\frac{\lambda^2 n(2n^2-3n+1)}{12} - 0.108998\frac{\lambda^2 n(n-1)(n-2)}{3} - \frac{\lambda^2 n(n-1)(n-2)(n-3)}{8\pi}.$$

Remark 5.3.10. If $\mathbf{X}$ is exchangeable and $\mathbf{N} \in \mathcal{S}_{n,p},\ p \geq 1$ in Theorem 5.3.6, then $P_e(\phi_{\text{lin}}, \mathcal{K}) = P_e(\phi_{\text{opt}}, \mathcal{K})$. This follows since under these conditions, from Theorem 5.3.3 we have $\phi_{\text{opt}}(\cdot) = \phi_{\text{lin}}(\cdot)$.

5.3.3 First-Order Approximation for P_e

From Section 5.3.2, one can infer that the private ranking recovery problem is noise dominated, i.e., the error probability is large even when σ is small. For instance, Example 5.3.8 and Example 5.3.9 suggest that the first-order coefficient c_1 grows quadratically with n. Thus, it becomes important to analyze the problem in the low-noise regime, where a reliable permutation recovery can be possible (i.e., $P_e \ll 1$). Towards this end, we next derive the first-order expansion of $P_e(\phi_{\text{lin}}, \mathcal{K})$ with respect to σ for any exchangeable $\mathbf{N}$ (note that Corollary 5.3.7 assumed $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$). The proof of the corollary below can be found in Appendix C.6.

Corollary 5.3.11. *Let* $\mathbf{N}$ *be exchangeable and* $V = N_1 - N_2$. *Assume that* $f_{W_i}(w) < \infty,\ \forall w$. *Then, in the low-noise regime, the first-order approximation of* P_e *is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \frac{C_{\mathbf{X}} \mathbb{E}\left[|V|\right]}{2}\sigma + O(\sigma^2), \tag{5.19}$$

with

$$C_{\mathbf{X}} = \sum_{i=1}^{n-1} f_{W_i}(0^+). \tag{5.20}$$

Remark 5.3.12. The first-order approximation of $P_e(\phi_{\text{lin}}, \mathcal{K})$ in (5.19) decouples the effects of the input data distribution (captured by $C_{\mathbf{X}}$) and of the noise distribution (captured by $\mathbb{E}[|V|]$). The assumptions of Corollary 5.3.11 are not too restrictive: as shown in Section 4.4.3, $f_{W_i}(\cdot)$ is bounded if $\mathbf{X}$ is i.i.d., and the PDF of X is bounded.

Remark 5.3.13. For the expansion of $P_e(\phi_{\text{lin}}, \mathcal{K})$ in (5.19), a natural question arises: How accurate is this? Figure 5.2 (see more figures in Appendix C.5.1) shows that this approximation

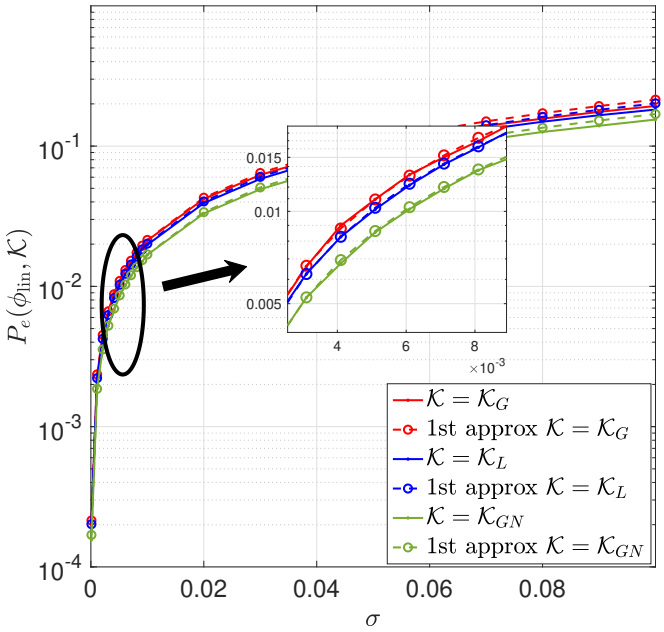


Figure 5.2: $P_e(\phi_{\text{lin}}, \mathcal{K})$ vs. its first-order approximation.

is indeed accurate when $\mathbf{N} \in \mathbb{R}^n$ is i.i.d. according to three different distributions, namely Gaussian (red curve), Laplace (blue curve), and generalized normal with $p = 0.5$ (green curve). Our rationale for choosing such distributions is because in Section 5.4, we will establish privacy-utility trade-offs for them. In Figure 5.2, the components of $\mathbf{X}$ were chosen to be i.i.d. according to $\text{Unif}(0, 100)$ with $n = 20$. The solid curves (probability of error) were obtained by Monte-Carlo simulation with 10^6 iterations, while the dashed curves (first order approximation of the error probability) where obtained by simply evaluating (5.19).

5.3.4 Input Data Vector with i.i.d. Entries

We here show that, for i.i.d. $X_i \sim X$, the dependence of the first-order approximation of $P_e(\phi_{\text{lin}}, \mathcal{K})$ in (5.19) on the distribution of $\mathbf{X}$ is rather weak (i.e., it only needs the L_2 norm of the PDF of X, and not the exact distribution). The approximation of $P_e(\phi_{\text{lin}}, \mathcal{K})$ in (5.19) depends on the distribution of $\mathbf{X}$ only through $C_{\mathbf{X}}$, and this term can be expressed in closed-form as stated in the following proposition (proof in Appendix C.7).

Proposition 5.3.14. *Let* $\mathbf{X}$ *consist of i.i.d. random variables with PDF* $f_X(\cdot)$*. Then,*

$$C_{\mathbf{X}} = n(n-1)\|f_X\|_2^2, \text{ where } \|f_X\|_2 = \sqrt{\int_{-\infty}^{\infty} f_X^2(x)\mathrm{d}x}. \tag{5.21}$$

According to Proposition 5.3.14 the first-order approximation of $P_e(\phi_{\text{lin}}, \mathcal{K})$ depends on the distribution of $\mathbf{X}$ only through the L_2 norm of its PDF. The significance of this result is that we do not need to know the exact distribution of the data vector to analyze $P_e(\phi_{\text{lin}}, \mathcal{K})$.

Remark 5.3.15. Proposition 5.3.14 shows that $C_{\mathbf{X}}$ grows quadratically in n. We now provide a few evaluations of $C_{\mathbf{X}}$ in (5.21):

- If $X \sim \text{Unif}(a, b)$, $C_{\mathbf{X}} = \frac{n(n-1)}{b-a}$;
- If $X \sim \text{Exp}(\lambda)$, $C_{\mathbf{X}} = \frac{\lambda n(n-1)}{2}$;
- If $X \sim \mathcal{N}(0, 1)$, $C_{\mathbf{X}} = \frac{n(n-1)}{2\sqrt{\pi}}$.

5.4 Privacy and Utility Trade-off

In this section, we investigate the relationship between privacy (measured by the (α, ϵ)-RDP in Definition 5.2.2) and utility measured by $P_e(\phi_{\text{lin}}, \mathcal{K})$. In particular, we focus on the low-noise regime where, as highlighted in Section 5.3.3, a reliable permutation recovery is possible.

For a proper definition of DP, we need to consider "well-behaved" query functions [16]. This is the so-called sensitivity property which, for a query function $q(\cdot)$, requires that the sensitivity (formally defined below) is finite.

Definition 5.4.1 (ℓ_p sensitivity [77])**.** For all $(\mathbf{X}, \tilde{\mathbf{X}}) \in \mathcal{X}^2$ such that $d_H(\mathbf{X}, \tilde{\mathbf{X}}) \leq 1$, the ℓ_p sensitivity of a query q is defined as

$$\Delta_p(q) = \max_{(\mathbf{X}, \tilde{\mathbf{X}}) \in \mathcal{X}^2 : d_H(\mathbf{X}, \tilde{\mathbf{X}}) \leq 1} \|q(\mathbf{X}) - q(\tilde{\mathbf{X}})\|_p, \tag{5.22}$$

where $p > 0$.

The ℓ_p sensitivity is a generalized version of the ℓ_1 sensitivity for the Laplace mechanism [67] and of the ℓ_2 sensitivity for the Gaussian mechanism [78]. In our framework, we have that the query function $q(\cdot)$ is the identity function, i.e., $q(\mathbf{x}) = \mathbf{x}$. Thus, in order to have

a finite ℓ_p sensitivity in (5.22), we need to have a domain constraint on the data input, namely $\mathbf{X} \in \mathcal{X}$ where $\mathcal{X} = \{\mathbf{x} : \mathbf{x} \in [0, \ell]^n\}$. With this, from (5.22), we have that the ℓ_p sensitivity is given by

$$\Delta_p(q) = \Delta(\mathbf{X}) = \max_{(\mathbf{X},\tilde{\mathbf{X}})\in\mathcal{X}^2 : d_H(\mathbf{X},\tilde{\mathbf{X}})\leq 1} \|\mathbf{X} - \tilde{\mathbf{X}}\|_p = \ell, \ \forall p > 0, \tag{5.23}$$

and is finite. In what follows, we let $\Delta(\mathbf{X}) = \ell$ denote the ℓ_p sensitivity for any $p > 0$, i.e., this notation indicates that the ℓ_p sensitivity is independent of $p > 0$. Next, in Section 5.4.1, we derive a general expression for the privacy-utility trade-off, which holds for any additive noise mechanism. Then, we evaluate it for practically relevant additive noise mechanisms, such as the Laplace (Section 5.4.2), the Gaussian (Section 5.4.3), and the generalized normal (Section 5.4.4) mechanisms.

5.4.1 On the General Trade-off

For the additive noise mechanism in (5.2), given (α, ϵ) and the sensitivity $\Delta(\mathbf{X}) = \ell$ in (5.23), we define the following operation,

$$\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell) = \inf\{\sigma : \mathsf{RDP}_\alpha\left(\mathcal{K}(\sigma, \mathbf{X})\right) \leq \epsilon, \Delta(\mathbf{X}) = \ell\}. \tag{5.24}$$

In words, $\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell)$ is the smallest standard deviation of $\mathcal{K}(\cdot, \cdot)$ that ensures that we meet the (α, ϵ)-RDP constraint when the query sensitivity is equal to ℓ. If the set in (5.24) is empty, then we set $\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell) = \infty$.

With the definition in (5.24) in mind and by using the first-order expansion of $P_e(\phi_{\text{lin}}, \mathcal{K})$ in Corollary 5.3.11, we arrive at the following general privacy-utility trade-off.

Proposition 5.4.2. *Consider an additive noise mechanism $\mathcal{K}(\sigma, \mathbf{X})$ as in* (5.2) *that adopts $\mathbf{N} \in \mathcal{S}_{n,p}$. Let the assumptions in Corollary 5.3.11 hold. Then, the privacy-utility trade-off for the ranking recovery problem is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \frac{\mathbb{E}[|V|]C_{\mathbf{X}}}{2}\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell) + O\left(\left(\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell)\right)^2\right), \tag{5.25}$$

where V and $C_{\mathbf{X}}$ are defined in Corollary 5.3.11.

Remark 5.4.3. The Taylor series of the error probability in Theorem 5.3.6 allows to characterize

higher order approximations for $P_e(\phi_{\text{lin}}, \mathcal{K})$, which in principle can lead to more accurate trade-offs in (5.25). The error term $O((\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell))^2)$ in (5.25) arises from the approximation error in the Taylor expansion of P_e.

The expression in (5.25) can, in principle, be used to find a privacy-utility trade-off for *any* additive noise mechanism. As expected, from (5.25) we note that $P_e(\phi_{\text{lin}}, \mathcal{K})$: (i) decreases as ϵ increases, and (ii) increases with the data size n. Moreover, for i.i.d. data, by using the closed-form expression in (5.21), we obtain the following trade-off,

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \frac{n(n-1)\mathbb{E}\left[|V|\right]\|f_X\|_2^2}{2}\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell) + O\left(\left(\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)\right)^2\right). \tag{5.26}$$

In the rest of this section, we seek to evaluate Proposition 5.4.2 and provide results in terms of ϵ instead of the implicit function $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$. Towards this end, we consider several important mechanisms for which the behavior of $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$ can be determined as a function of ϵ and ℓ. For some mechanisms, the expression for $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$ is already known and simply needs to be remapped to our notation (e.g., Gaussian mechanism). For other mechanisms, the expression for ϵ exists in closed-form, but the inverse $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$ does not have a closed-form (e.g., Laplace mechanism). In such a case, we provide upper and lower bounds on $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$ that indicate its behavior. Yet, in other cases, we find new expressions for $\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell)$ (e.g., generalized normal mechanisms for $\alpha = \infty$).

5.4.2 Laplace Mechanism

We consider a randomized mechanism $\mathcal{K}_L(\sigma, \mathbf{X})$ that consists of adding Laplace noise. Such a mechanism gives (α,ϵ)-RDP as shown in the next result, the proof of which uses the results by [79] (proof in Appendix C.8).

Proposition 5.4.4. *For $\alpha > 1$, the randomized mechanism $\mathcal{K}_L(\sigma,\mathbf{X})$ in* (5.2) *with $\mathbf{N}$ being i.i.d. according to Lap$(0,b)$ gives (α,ϵ)-RDP with ϵ given by*

$$\epsilon = \frac{1}{\alpha-1}\ln\frac{\alpha e^{-(1-\alpha)\ell/(\sigma b)} - (1-\alpha)e^{-\alpha\ell/(\sigma b)}}{2\alpha-1}. \tag{5.27}$$

Moreover, letting $c_\alpha = \frac{1}{\alpha-1}\ln\frac{\alpha}{2\alpha-1}$, we have that

$$\frac{\ell}{\sigma b} + c_\alpha \le \epsilon \le \frac{\ell}{\sigma b} + c_\alpha + \frac{1}{\alpha}e^{-\frac{(2\alpha-1)\ell}{\sigma b}}. \tag{5.28}$$

We note that Proposition 5.4.4 is a generalization of the RDP analysis for the Laplace mechanism in [17] that considered the 1-dimensional case. Although the generalization under the i.i.d. assumption is straightforward and follows a similar proof, we here reported the proof of Proposition 5.4.4 for completeness. In addition, the upper and lower bounds on ϵ are also provided, which we leverage next to provide the privacy-utility trade-off. Furthremore, the gap (i.e., difference) between the upper bound and the lower bound in (5.28) is given by $\frac{1}{\alpha}e^{-\frac{(2\alpha-1)\ell}{\sigma b}}$. Thus, we can conclude that the bounds in (5.28) are moderately tight when α is not too small, and the bounds become tight as $\alpha \to \infty$.

We now combine Proposition 5.4.4 and Corollary 5.4.2 and obtain an explicit first-order approximation of P_e in terms of ϵ and ℓ for the Laplace mechanism in the following corollary (proof in Appendix C.9).[4]

Corollary 5.4.5. *Let $\mathcal{K}_L(\sigma, \mathbf{X})$ be such that $\mathbf{N}$ is i.i.d. according to Lap $\left(0, \frac{1}{\sqrt{2}}\right)$. Let the assumptions in Corollary 5.3.11 hold. Then, for $\alpha > 1$, the privacy-utility trade-off is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}_L) = \frac{3C_{\mathbf{X}}}{4\sqrt{2}}\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell) + O\left(\frac{1}{\epsilon^2}\right), \tag{5.29}$$

where

$$\frac{\sqrt{2}\ell}{\left(\epsilon + \frac{1}{\alpha-1}\ln\frac{2\alpha-1}{\alpha}\right)} \leq \mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell) \leq \frac{\sqrt{2}\ell}{\epsilon}. \tag{5.30}$$

Although Corollary 5.4.5 provides the trade-off in terms of upper and lower bounds, it implies that the trade-off is at least $P_e \propto \frac{1}{\epsilon}$ by considering the lower bound on $\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell)$. We note that for the case of $\alpha = \infty$, which is equivalent to ϵ-DP, the bound in (5.30) becomes exact and $\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell) = \frac{\sqrt{2}\ell}{\epsilon}$.

Remark 5.4.6. Since $\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell)$ does not have a closed-form, the bounds on $\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell)$ in (5.30) were provided to indicate its behavior with respect to ϵ and ℓ. However, if one needs an exact value of $\mathsf{RDP}_{\alpha}^{-1}(\epsilon, \ell)$ for a given (ϵ, ℓ), this can easily be done numerically by inverting (5.27).

As an example, we next evaluate (5.29) when $\mathbf{X}$ is i.i.d. and has a uniform distribution.

Example 5.4.7. If $\mathbf{X} \sim \text{Unif}([0, \ell]^n)$, then $C_{\mathbf{X}} = \frac{n(n-1)}{\ell}$ and the trade-off in Corollary 5.4.5 becomes

$$P_e(\phi_{\text{lin}}, \mathcal{K}_L) = \frac{3n(n-1)}{4}\mathsf{R}_{\alpha}^{-1}(\epsilon) + O\left(\frac{1}{\epsilon^2}\right),$$

[4]The approximation arises from the Taylor series of P_e in Corollary 5.3.11.

where

$$\frac{1}{\epsilon + \frac{1}{\alpha-1}\ln\frac{2\alpha-1}{\alpha}} \leq \mathsf{R}_\alpha^{-1}(\epsilon) \leq \frac{1}{\epsilon}.$$

5.4.3 Gaussian Mechanism

We here analyze a mechanism $\mathcal{K}_G(\sigma, \mathbf{X})$ that consists of adding Gaussian noise. This gives (α, ϵ)-RDP as shown in the next result, the proof of which can be found in Appendix C.10 and uses the results in [79]. We note that this result was already derived by [17], but we report it here for completeness.

Proposition 5.4.8. *$\mathcal{K}_G(\sigma, \mathbf{X})$ in* (5.2) *with $\mathbf{N}$ being i.i.d. according to $\mathcal{N}(0,1)$ gives $\left(\alpha, \frac{\alpha\ell^2}{2\sigma^2}\right)$-RDP. Consequently,*

$$\mathsf{RDP}_\alpha^{-1}(\epsilon, \ell) = \sqrt{\frac{\alpha\ell^2}{2\epsilon}}. \tag{5.31}$$

We now evaluate the trade-off stated in Proposition 5.4.2. For independent standard Gaussian random variables N_1 and N_2, we have that $\mathbb{E}[|V|] = \mathbb{E}[|N_1 - N_2|] = \frac{2}{\sqrt{\pi}}$. By leveraging Proposition 5.4.8 and Corollary 5.3.11, we then obtain the privacy-utility trade-off for the Gaussian mechanism as shown in the following corollary.

Corollary 5.4.9. *Consider the Gaussian mechanism $\mathcal{K}_G(\sigma, \mathbf{X})$ with $\mathbf{N}$ being i.i.d. according to $\mathcal{N}(0,1)$. Let the assumptions in Corollary 5.3.11 hold. Then, for $\alpha \geq 1$, the privacy-utility trade-off is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}_G) = \frac{\ell C_{\mathbf{X}}}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{\epsilon}} + O\left(\frac{1}{\epsilon}\right). \tag{5.32}$$

From (5.32) we observe that the Gaussian mechanism gives a P_e that is inversely proportional to $\sqrt{\epsilon}$, while the Laplace mechanism in (5.29) offers a P_e that scales inversely proportional to ϵ as shown in Corollary 5.4.5. Thus, we can conclude that the Laplace mechanism outperforms the Gaussian mechanism in terms of the rate of the privacy-utility trade-off. We complete this subsection by giving an example when $\mathbf{X}$ is i.i.d. and has a uniform distribution.

Example 5.4.10. If $\mathbf{X} \sim \text{Unif}([0,\ell]^n)$, then $C_{\mathbf{X}} = \frac{n(n-1)}{\ell}$ and the trade-off in Corollary 5.4.9 becomes

$$P_e(\phi_{\text{lin}}, \mathcal{K}_G) = \frac{n(n-1)}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{\epsilon}} + O\left(\frac{1}{\epsilon}\right).$$

5.4.4 Generalized Normal Mechanism

Corollary 5.4.5 and Corollary 5.4.9 suggest that $P_e \propto (1/\epsilon)^{1/p}$, where p is the power of the exponent in the noise PDF. In other words, the smaller the p is, the better the trade-off appears to be. Motivated by this observation, we consider a generalized normal mechanism [77] denoted by $\mathcal{K}_{GN}$ where $\mathbf{N}$ is i.i.d. according to $\mathcal{GN}(0, a, p)$ with $p \leq 1$. Although p can be greater than 1, we only consider $p \leq 1$ as motivated by the trade-off $P_e \propto (1/\epsilon)^{1/p}$. Different from the previous RDP analysis for the Laplace and Gaussian mechanisms, we here study only ϵ-DP for $\mathcal{K}_{GN}$ (i.e., $\alpha = \infty$). Recall that ϵ-DP offers a stronger privacy guarantee than RDP. The ϵ-DP of $\mathcal{K}_{GN}$ is given in the next proposition (proof in Appendix C.11).

Proposition 5.4.11. *Let* $\mathbf{N}$ *be i.i.d. according to* $N \sim \mathcal{GN}(0, h(p), p)$ *with* $p \leq 1$ *and* $h(p) = \sqrt{\frac{\Gamma(p^{-1})}{\Gamma(3p^{-1})}}$, *where* $\Gamma(\cdot)$ *is the gamma function. Then, the generalized normal mechanism* $\mathcal{K}_{GN}(\sigma)$ *gives* ϵ*-DP with*

$$\epsilon = \left(\frac{\ell}{\sigma h(p)}\right)^p. \tag{5.33}$$

Consequently,

$$\mathsf{RDP}_\infty^{-1}(\epsilon, \ell) = \frac{\ell}{h(p)} \left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}. \tag{5.34}$$

Remark 5.4.12. We note that the work of [77] only considered integer values for the parameter p. The above result extends the work of [77] to any $p \in (0, 1]$.

We combine Proposition 5.4.2 and Proposition 5.4.11 and obtain the trade-off in the corollary below.

Corollary 5.4.13. *Consider the generalized normal mechanism* $\mathcal{K}_{GN}(\sigma, \mathbf{X})$ *with* $p \leq 1$. *Let the assumptions in Corollary 5.3.11 hold. Then, the privacy-utility trade-off is given by*

$$P_e(\phi_{\text{lin}}, \mathcal{K}_{GN}) = \frac{\mathbb{E}[|N - N'|]\ell C_{\mathbf{X}}}{2h(p)} \left(\frac{1}{\epsilon}\right)^{\frac{1}{p}} + O\left(\frac{1}{\epsilon^{\frac{2}{p}}}\right), \tag{5.35}$$

where N*and* N' *are independent and* $N' \overset{d}{=} N$.

Remark 5.4.14. Corollary 5.4.13 confirms our observation that the smaller the p is, the better the trade-off is. Thus, for $\alpha = \infty$ (or ϵ-DP), the generalized normal distribution with $p \leq 1$ offers a better privacy-utility trade-off than the Laplace and Gaussian mechanisms. In addition,

note that the constant in the first-order term of (5.35) can be upper bound by using Jensen's inequality as follows,

$$\mathbb{E}[|N - N'|] \leq \sqrt{\mathbb{E}[|N - N'|^2]} = \sqrt{2\mathrm{Var}(N)} = \sqrt{2},$$

where we have used the fact that $\mathrm{Var}(N) = 1$. Furthermore, we seek to minimize (5.35) with respect to p in order to find the best generalized normal mechanism given an ϵ-DP constraint. We refer to Appendix C.12, where we discuss this and provide the best p.

5.5 Conclusions

We studied the private ranking recovery problem within the DP framework. We designed a low-complexity decoder and characterized sufficient conditions for its optimality. We derived the Taylor series of the error probability when such a decoder is used, as well as the first-order approximation of it. We leveraged the first-order approximation of the error probability, along with the (α, ϵ)-RDP, to obtain utility-privacy trade-offs for the Gaussian, Laplace, and generalized normal mechanisms. These results allow us to compare different noise mechanisms in order to determine the best utility-privacy trade-off. In addition, our results show that the problem of private ranking recovery is noise dominated, i.e., the error probability is large even for small values of the noise variance. This suggests that the exact recovery imposed in our work might need to be relaxed. Finally, possible future directions include the following: (i) *partial recovery* in which we seek to recover the permutation of only part of the input data; (ii) *approximate recovery* in which we allow a fixed number of errors given a ranking distance function [64] (e.g., Hamming distance, Kendall's tau distance); (iii) investigating or generalizing the results in this paper to hold universally, for any distribution on the input data vector.

Chapter 6

Approximate Permutation Recovery

In this chapter, we study the problem of permutation recovery up to a distortion, so-called approximate permutation recovery. Specifically, we seek to estimate the true permutation of a data vector up to certain estimation error. We formulate the problem using several ranking distances, and we proved that the decoders used in the previous chapters have similar properties, such as optimality. We characterize the error probability of the approximate permutation recovery and compare it with the one of exact permutation recovery studied in previous chapters.

6.1 Introduction

Today, ranking data is a pervading task in several applications, such as search engines [80], biomedical [81], recommender systems [82], feature matching [2], and communication systems [83]. However, the data might be *noisy*, e.g., because of privacy considerations [78], 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 chapter, we introduce an *approximate* version of the ranking recovery problem previously explored. Specifically, the challenge addressed in earlier sections involved the retrieval of the *exact* sequence (or ranking) in which an input dataset was organized prior to being disrupted by additive noise. In this chapter, we adopt a more flexible stance on the problem, permitting a certain degree of distortion in the ranking estimation (refer to Section 6.2). Our focus is primarily on scenarios when the data is perturbed by isotropic Gaussian noise, and we quantify the

distortion through a distance metric between the estimated and actual rankings of the initial data vector. We initially demonstrate (refer to Section 6.3) that the optimal solution for this problem, in terms of minimizing error probability, is attained through the *linear decoder* outlined in prior chapters. This decoder, known for its simplicity in just yielding the observed noisy ranking, is celebrated for its computational efficiency, scaling polynomially with data dimensionality. Subsequently, our analysis moves into the error probability incurred with the linear decoder in the low-noise regime (see Section 6.4). Here, it is revealed that the error probability increases sub-linearly with the noise's standard deviation σ, marking a distinct departure from the linear relationship observed in the exact recovery scenarios discussed in Chapter 4 and Chapter 5. Such findings highlights the lesser noise susceptibility of this approximate ranking recovery. These conclusions are drawn under the presumption of mild conditions on the distance metric, applicable to commonly used measures such as the Hamming distance and Kendall's tau rank distance.

6.2 Notation and Problem Formulation

In this section we provide several definitions that are required to formally define ranking and distance function between two rankings.

Definition 6.2.1 (Permutation)**.** We denote by $\pi_{\mathbf{x}}$ the permutation of $\mathbf{x} \in \mathbb{R}^n$ such that

$$(\mathbf{x}_{\pi_{\mathbf{x}}})_1 \leq ... \leq (\mathbf{x}_{\pi_{\mathbf{x}}})_i \leq ... \leq (\mathbf{x}_{\pi_{\mathbf{x}}})_n, \tag{6.1}$$

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

Example 6.2.2. If $\mathbf{x} = (-2, 3, -6, 1, 2)$, then we have

$$\pi_{\mathbf{x}} = (3, 1, 4, 5, 2), \text{ and } \mathbf{r}_{\mathbf{x}} = (2, 5, 1, 3, 4). \tag{6.2}$$

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

Definition 6.2.3 (Hamming distance)**.** 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

$$\mathsf{d}_H(\mathbf{r_u}, \mathbf{r_v}) = |\{i : (\mathbf{r_u})_i \neq (\mathbf{r_v})_i\}| = \sum_{i=1}^{n} \mathbb{1}\{(\mathbf{r_u})_i \neq (\mathbf{r_v})_i\}. \tag{6.3}$$

Definition 6.2.4 (Kendall's Tau rank distance)**.** 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

$$\mathsf{d}_K(\mathbf{r_u}, \mathbf{r_v}) = |\{(i,j) : i < j, \mathrm{sgn}((\mathbf{r_u})_i - (\mathbf{r_u})_j) \neq \mathrm{sgn}((\mathbf{r_v})_i - (\mathbf{r_v})_j)\}|, \tag{6.4}$$

where sgn denotes the sign function.

Example 6.2.5. 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

$$\begin{aligned}\mathbf{r_y} \in \{&(1,3,4,2), (1,4,2,3), (1,2,3,4),\\ &(3,1,2,4), (2,3,1,4), (4,3,2,1)\},\end{aligned}$$

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

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

6.2.1 Preliminaries and Known Results

We consider the following model,

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{6.5}$$

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 section, 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 (6.5). In the previous chapters, we considered this problem under an *exact* ranking recovery constraint, i.e., we were interested in recovering the

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

exact ranking according to which the input data vector $\mathbf{X}$ was sorted.[3] In particular, we showed that the decision rule, referred to as decoder, 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 the previous chapters, we showed that, for any value of the noise standard deviation σ, for the *exact* ranking recovery we have that

$$\phi_{\text{lin}} \in \operatorname*{argmin}_{\phi} P_e(\phi, \sigma) = \operatorname*{argmin}_{\phi} \Pr(\phi(\mathbf{Y}) \neq \mathbf{r_X}), \tag{6.6}$$

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. 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.

6.2.2 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 6.2.3, Kendall's tau in Definition 6.2.4) 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.

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

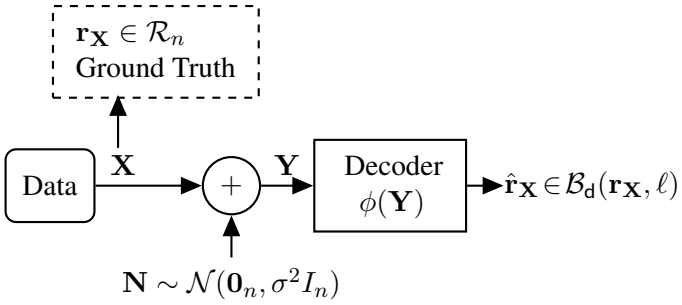


Figure 6.1: Graphical representation of the approximate ranking recovery.

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

$$\mathcal{B}_\mathsf{d}(\mathbf{r_c}, \ell) = \{\mathbf{r_x} \in \mathcal{R}_n : \mathsf{d}(\mathbf{r_c}, \mathbf{r_x}) \leq \ell\}, \tag{6.7}$$

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

Example 6.2.6. Consider the case $n = 4$, for which $|\mathcal{R}_4| = 24$. Let d be the Hamming distance in Definition 6.2.3, and $\ell = 2$. For $\mathbf{r_c} = (1, 2, 3, 4)$, we have that

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

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. 6.1 provides a graphical representation of the *approximate* ranking recovery problem. We note that due to the assumption **A1**, setting $\ell = 0$ in $\mathcal{B}_\mathsf{d}(\mathbf{r_x}, \ell)$ recovers the *exact* version of the problem studied in the previous chapters.

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

$$P_e(\phi, \mathsf{d}, \ell) = \Pr(\mathsf{d}(\mathbf{r_X}, \phi(\mathbf{Y})) > \ell). \tag{6.8}$$

In particular, our focus will be on characterizing an optimal (i.e., that minimizes (6.8)) decoder (see Section 6.3), and on understanding how (6.8) varies with respect to the noise standard deviation $\sigma \in \mathbb{R}_+$ (see Section 6.4).

6.3 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 (6.8) in the estimation task. With the definition in (6.8), an optimal decoder ϕ_{opt} is given by

$$\phi_{\text{opt}} \in \operatorname*{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell), \tag{6.9}$$

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

$$\begin{aligned} P_c(\phi, \mathsf{d}, \ell) &= \Pr(\mathsf{d}(\mathbf{r}_{\mathbf{X}}, \phi(\mathbf{Y})) \leq \ell) \\ &= \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\mathbf{r}_{\mathbf{X}}, \ell)). \end{aligned} \tag{6.10}$$

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

Lemma 6.3.1. *If, for all $\mathbf{y} \in \mathbb{R}^n$, a decoder $\hat{\phi} : \mathbb{R}^n \to \mathcal{R}_n$ satisfies*

$$\hat{\phi}(\mathbf{y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell), \quad \tau = \operatorname*{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y}), \tag{6.11}$$

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 (6.10) as

$$\begin{aligned} P_c(\phi, \mathsf{d}, \ell) &= \sum_{\tau \in \mathcal{R}_n} \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell), \mathbf{r}_{\mathbf{X}} = \tau) \\ &= \sum_{\tau \in \mathcal{R}_n} \Pr(\phi(\mathbf{Y}) \in \mathcal{B}_{\mathsf{d}}(\tau, \ell) \mid \mathbf{r}_{\mathbf{X}} = \tau)\, p_{\mathbf{r}_{\mathbf{X}}}(\tau) \\ &= \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \Pr(\phi(\mathbf{Y}) = \omega \mid \mathbf{r}_{\mathbf{X}} = \tau)\, p_{\mathbf{r}_{\mathbf{X}}}(\tau) \\ &\overset{(a)}{=} \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \int_{\mathbf{y} \in \mathcal{D}_\omega} f_{\mathbf{Y}|\mathbf{r}_{\mathbf{X}}}(\mathbf{y}|\tau)\, p_{\mathbf{r}_{\mathbf{X}}}(\tau)\, \mathrm{d}\mathbf{y} \\ &\overset{(b)}{=} \sum_{\tau \in \mathcal{R}_n} \sum_{\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)} \int_{\mathbf{y} \in \mathcal{D}_\omega} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\tau|\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\, \mathrm{d}\mathbf{y}, \end{aligned} \tag{6.12}$$

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 (6.9), we need to maximize $P_c(\phi, \mathsf{d}, \ell)$ with respect to ϕ. Equivalently, with reference to (6.12), 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 (6.12). 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 (6.12). 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 $\tau = \mathrm{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r}_{\mathbf{X}}|\mathbf{Y}}(\eta|\mathbf{y})$, we assign $\mathbf{y}$ to $\mathcal{D}_\omega$, where $\omega \in \mathcal{B}_{\mathsf{d}}(\tau, \ell)$. This concludes the proof of Lemma 6.3.1. □

By leveraging Lemma 6.3.1, we are now ready to prove our first main result, which shows that the linear decoder $\phi_{\mathrm{lin}}(\mathbf{y}) = \mathbf{r}_{\mathbf{y}}$ is indeed optimal for the approximate recovery problem.

Theorem 6.3.2. *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*

$$\phi_{\mathrm{lin}} \in \operatorname*{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell). \tag{6.13}$$

Proof. We consider the Maximum a Posteriori (MAP) decision rule [54], i.e.,[4]

$$\phi_{\mathrm{MAP}}(\mathbf{y}) = \operatorname*{argmax}_{\eta \in \mathcal{R}_n} p_{\mathbf{r_X}|\mathbf{Y}}(\eta|\mathbf{y}). \tag{6.14}$$

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 (6.11) in Lemma 6.3.1 are satisfied and hence, it is guaranteed that $\phi_{\mathrm{MAP}} \in \operatorname{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell)$, i.e., ϕ_{MAP} is an optimal decoder. In the previous chapters, 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 $\phi_{\mathrm{lin}} = \phi_{\mathrm{MAP}}$. This readily implies that $\phi_{\mathrm{lin}} \in \operatorname{argmax}_{\phi} P_c(\phi, \mathsf{d}, \ell)$, and concludes the proof of Theorem 6.3.2. □

Remark 6.3.3. The assumption **A1** in Theorem 6.3.2 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 6.2.3 and the Kendall's tau rank distance in Definition 6.2.4 satisfy these conditions.

6.4 $P_e(\phi_{\mathrm{lin}}, \mathsf{d}, \ell)$ versus σ

Theorem 6.3.2 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 $P_e(\phi_{\mathrm{lin}}, \mathsf{d}, \ell)$, 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 of error is linear in σ), we show that for the *approximate* version of the problem $P_e(\phi_{\mathrm{lin}}, \mathsf{d}, \ell)$ exhibits a sublinear behavior in σ in the low-noise regime (see Theorem 6.4.3). This result is also shown in Fig. 6.2 (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 on error probability of the approximate ranking recovery.

[4] We note that $\phi_{\mathrm{MAP}}(\mathbf{y})$ in (6.14) might not be unique; if this is the case, then we randomly select one of these possible choices.

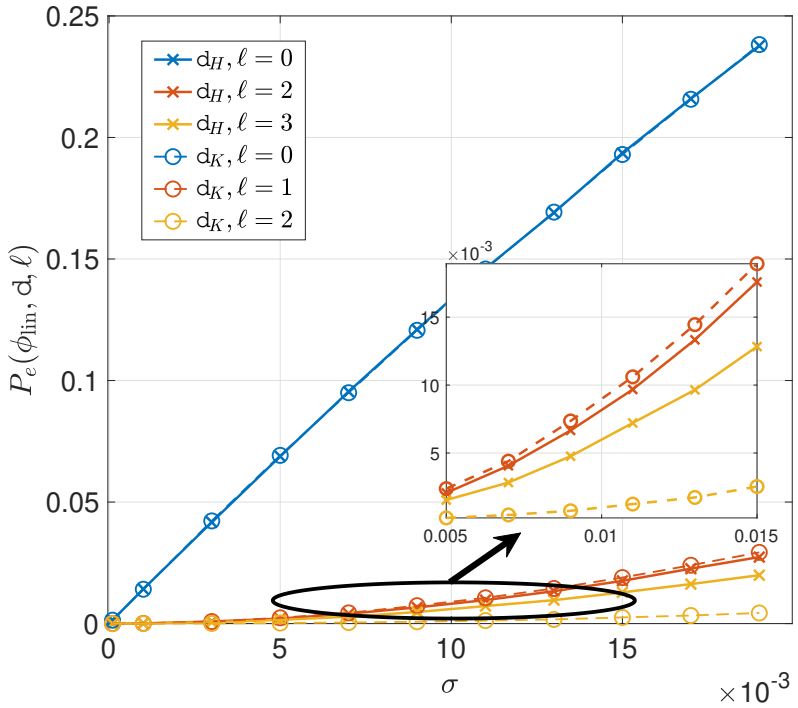


Figure 6.2: $P_e(\phi_{\text{lin}}, \mathsf{d}, \ell)$ in (6.8) versus σ with $\mathsf{d} \in \{\mathsf{d}_H, \mathsf{d}_K\}$ and $\ell \in \{0, 1, 2, 3\}$. We set $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_{10}, I_{10})$ and $\mathcal{N} \sim \mathcal{N}(\mathbf{0}_{10}, \sigma^2 I_{10})$.

Definition 6.4.1. Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector. The i-th order statistics [53] of $\mathbf{X}$ (i.e., the i-th smallest value of $\mathbf{X}$) satisfies

$$X_{i:n} : X_{1:n} \leq ... \leq X_{i:n} \leq ... \leq X_{n:n}. \tag{6.15}$$

Then, we say that the i-th spacing [60] of $\mathbf{X}$ is

$$W_i = X_{i+1:n} - X_{i:n}. \tag{6.16}$$

We now state the following lemma, the proof of which can be found in Section 6.5.

Lemma 6.4.2. *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* (6.16)*. Then,*

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

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

By leveraging Lemma 6.4.2, we can now prove the following theorem, which is the second main result of the paper.

Theorem 6.4.3. *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* (6.16)*. Consider a distance function* d *satisfying the assumptions* **A1** *and* **A2***, and let*

$$\mathsf{d}(\tau, P^{(i,i+1)}\tau) = \beta_i,\ \tau = (1, ..., n),\ \forall i \in [1 : n-1]. \tag{6.17}$$

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

$$\lim_{\sigma \to 0} \frac{1}{\sigma} P_e(\phi_{\text{lin}}, \mathsf{d}, \ell) = 0. \tag{6.18}$$

Proof. We start by observing that

$$\begin{aligned}
\Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k) &= \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \eta) \\
&\overset{(a)}{=} \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(P_{\tau,\eta}\mathbf{r_X}, P_{\tau,\eta}\mathbf{r_Y}) = k, P_{\tau,\eta}\mathbf{r_X} = \eta) \\
&\overset{(b)}{=} \sum_{\eta \in \mathcal{R}_n} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \tau) \\
&= n! \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k, \mathbf{r_X} = \tau) \\
&= \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau),
\end{aligned} \tag{6.19}$$

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 (6.19) we then obtain

$$\begin{aligned}
P_e(\phi_{\text{lin}}, \mathsf{d}, 0) &= \sum_{k>0} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k) \\
&= \sum_{k>0} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&\overset{(a)}{=} \sum_{0<k\leq\beta^\star} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&\quad + \sum_{k>\beta^\star} \Pr(\mathsf{d}(\mathbf{r_X}, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) \\
&= \sum_{0<k\leq\beta^\star} \Pr(\mathsf{d}(\tau, \mathbf{r_Y}) = k \mid \mathbf{r_X} = \tau) + P_e(\phi, \mathsf{d}, \beta^\star) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n-1} \Pr\left(\mathbf{r_Y} = P^{(i,i+1)}\tau \;\middle|\; \mathbf{r_X} = \tau\right) + P_e(\phi, \mathsf{d}, \beta^\star), \qquad (6.20)
\end{aligned}$$

where (a) follows by letting $\beta^\star = \max_{i\in[1:n-1]}\{\beta_i\}$, and (b) is due to (6.17).

From Corollary 4.4.3, we know that

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

and from Lemma 6.4.2, 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 (6.20), imply that $\lim_{\sigma\to 0} \frac{1}{\sigma} P_e(\phi_{\text{lin}}, \mathsf{d}, \beta^\star) = 0$. We conclude the proof of Theorem 6.4.3 by noting that for any $\ell \geq \beta^\star$, we have that $P_e(\phi_{\text{lin}}, \mathsf{d}, \ell) \leq P_e(\phi_{\text{lin}}, \mathsf{d}, \beta^\star)$, which implies $\lim_{\sigma\to 0} \frac{1}{\sigma} P_e(\phi_{\text{lin}}, \mathsf{d}, \ell) = 0$ for all $\ell \geq \beta^\star$. □

Remark 6.4.4. The $1/\sigma$ in Theorem 6.4.3 is used to prove the sublinear behavior of P_e in the low-noise regime (i.e., the limit in (6.18) is indeed zero). Theorem 6.4.3 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 $\mathcal{H}_\tau = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{r_x} = \tau\}$ 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 6.2.3 and Definition 6.2.4.

Corollary 6.4.5. *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 (6.18) in Theorem 6.4.3 holds. This concludes the proof of Corollary 6.4.5. □

Corollary 6.4.6. *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 6.4.3, for any $\ell \geq 1$ we have that (6.18) in Theorem 6.4.3 holds. This concludes the proof of Corollary 6.4.6. □

6.5 Proof of Lemma 6.4.2

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,

$$
\begin{aligned}
&\Pr\left(\mathbf{r}_{\mathbf{Y}} = P^{(i,i+1)}\tau \;\middle|\; \mathbf{r}_{\mathbf{X}} = \tau\right) \\
&\overset{(a)}{=} \Pr\left(\bigcap_{t\in\mathcal{I}_i}\{Y_t \leq Y_{t+1}\} \cap \mathcal{E}_i(\mathbf{Y}) \;\middle|\; \tau\right) \\
&\overset{(b)}{=} \Pr\left(\bigcap_{t\in\mathcal{I}_i}\{V_t \leq W_t\} \cap \mathcal{E}_i(\mathbf{X}+\mathbf{N}) \;\middle|\; \tau\right) \\
&\overset{(c)}{=} \Pr\left(\bigcap_{t\in\mathcal{I}_i}\{V_t \leq W_t\} \cap \tilde{\mathcal{E}}_i(\mathbf{X}+\mathbf{N}) \;\middle|\; \tau\right) \\
&\overset{(d)}{=} \Pr\left(V_i \leq -W_i\right) \Pr\left(\bigcap_{t\in\mathcal{I}_i}\{V_t \leq W_t\} \cap \mathcal{E}_i^{\star}(\mathbf{V},\mathbf{W}) \;\middle|\; V_i \leq -W_i\right),
\end{aligned} \tag{6.21}
$$

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 6.4.1 for which $W_t = X_{t+1} - X_t$ and letting $V_t = N_t - N_{t+1}$; note that, with reference to Definition 6.4.1 we have that $X_{i:n} \overset{d}{=} X_i$ 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,

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

where $(\mathrm{c1})$ follows by permuting N_i and N_{i+1}, and $(\mathrm{c2})$ follows since $\mathbf{r}_{\mathbf{X}} = \tau$ and by using Definition 6.4.1 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 (6.21). The first probability term in (6.21) is

$$\begin{aligned}\Pr\left(V_i \le -W_i\right) &= \Pr\left(Z \le -\frac{W_i}{\sigma\sqrt{2}}\right) \\ &= \int_0^\infty Q\left(\frac{w}{\sigma\sqrt{2}}\right) f_{W_i}(w)\,\mathrm{d}w \\ &= \int_0^\infty Q\left(u\right) f_{W_i}(\sqrt{2}\sigma u)\sqrt{2}\sigma\,\mathrm{d}u, \end{aligned} \tag{6.22}$$

where $Q(\cdot)$ is the standard Gaussian Q function, and the last equality follows by a change of variable. By dividing (6.22) by σ and taking $\sigma \to 0$, we obtain

$$\begin{aligned}\lim_{\sigma\to 0}\frac{\Pr\left(V_i \le -W_i\right)}{\sigma} &\overset{(a)}{=} \int_0^\infty Q\left(u\right) \lim_{\sigma\to 0} f_{W_i}(\sqrt{2}\sigma u)\sqrt{2}\,\mathrm{d}u \\ &= \int_0^\infty Q\left(u\right) f_{W_i}(0^+)\sqrt{2}\,\mathrm{d}u = \frac{f_{W_i}(0^+)}{\sqrt{\pi}}, \end{aligned} \tag{6.23}$$

where (a) follows from the dominated convergence theorem, which is verifiable since $f_{W_i}(w) \le \sup f_{W_i}(w) < \infty$, and $\int_0^\infty Q(u)\mathrm{d}u$ is integrable. The second probability term in (6.21) is

$$\begin{aligned}&\lim_{\sigma\to 0}\Pr\left(\bigcap_{t\in\mathcal{I}_i}\{V_t \le W_t\}\cap\mathcal{E}_i^\star(\mathbf{V},\mathbf{W})\middle| V_i \le -W_i\right) \\ &\overset{(a)}{=} \lim_{\sigma\to 0}\Pr\left(\bigcap_{t\in\mathcal{I}_i}\{\sigma\tilde{V}_t \le W_t\}\cap\mathcal{E}_i^\star(\sigma\tilde{\mathbf{V}},\mathbf{W})\middle| V_i \le -W_i\right) \overset{(b)}{=} 1, \end{aligned} \tag{6.24}$$

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 $\mathbf{W} \ge \mathbf{0}_{n-1}$. By using (6.21), (6.23) and (6.24), we obtain

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

and hence,

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

This concludes the proof of Lemma 6.4.2.

Chapter 7

Conclusion

In conclusion, this thesis systematically navigates the complex domain of permutation recovery from noisy observations, striking at the core of significant challenges in data science and privacy-preserving mechanisms. By meticulously addressing both exact and approximate permutation recovery under a spectrum of noise conditions, this work sheds light on the pivotal question: How to discern the original ordering of data from its noise-altered state?

The research begins by formulating the permutation recovery problem within a statistical hypothesis testing framework, and unveils the linear regime where optimal permutation estimation is achievable through a linear transformation of the noisy data, succeeded by sorting. This exploration, under Gaussian-distributed data and noise, delineates the critical influence of the noise covariance matrix, revealing that a flat spectrum with a limited number of distinct eigenvalues is essential for inducing the linear regime.

Delving into the error probabilities associated with linear decoders, this thesis uncovers the inherently noise-dominated landscape of permutation recovery. The detailed analysis across both low- and high-noise scenarios, and the characterizing error probability's scaling behavior, furnish profound insights into the intricate interplay between noise, data distribution, and recovery fidelity. The discovery that error probability scales linearly with the noise standard deviation, σ in the low-noise regime highlights the impact of noise for the permutation recovery problem.

Furthermore, advancing into the realm of data privacy, the thesis embarks on elucidating the delicate trade-off between estimation accuracy and privacy. Employing differential privacy metrics, such as ϵ-DP and (α, ϵ)-RDP, it evaluates the efficacy of diverse privacy-preserving

mechanisms, forging a path to optimize the trade-offs between privacy and the fidelity of ranking recovery.

In an attempt to generalize the problem, the thesis proposes an approximate version of ranking recovery and demonstrates that the probability of error exhibits sub-linear behavior in σ, in contrast to the linear behavior observed in exact recovery. This approach, characterized by its reduced susceptibility to noise, underscores the potential of approximate recovery techniques in enhancing the robustness and efficiency of permutation recovery processes.

The collective contributions of this thesis not only fortify the theoretical underpinnings of permutation recovery in noisy and privacy-sensitive environments but also lay down practical methodologies for addressing these challenges. As we look toward the horizon of data processing and privacy-preserving techniques, the findings and strategies delineated herein pave the way for future research, promising enhanced robustness, efficiency, and privacy in the face of ever-evolving data landscapes.

References

[1] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering structure of noisy data through hypothesis testing. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1307–1312, 2020.

[2] Minoh Jeong, Alex Dytso, Martina Cardone, and H. Vincent Poor. Recovering data permutations from noisy observations: The linear regime. *IEEE Journal on Selected Areas in Information Theory*, 1(3):854–869, 2020.

[3] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: High and low noise asymptotics. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1100–1105, 2021.

[4] Minoh Jeong, Alex Dytso, and Martina Cardone. Ranking recovery under privacy considerations. *Transactions on Machine Learning Research*, 2022.

[5] Minoh Jeong, Martina Cardone, and Alex Dytso. On the ranking recovery from noisy observations up to a distortion. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 1993–1998, 2022.

[6] Minoh Jeong, Alex Dytso, and Martina Cardone. Retrieving data permutations from noisy observations: Asymptotics. *IEEE Transactions on Information Theory*, 70(4):2999–3017, 2024.

[7] Minoh Jeong, Alex Dytso, and Martina Cardone. Gradient of error probability of M-ary hypothesis testing problems under multivariate Gaussian noise. *IEEE Signal Processing Letters*, 27:1909–1913, 2020.

[8] Minoh Jeong, Alex Dytso, and Martina Cardone. Functional properties of the Ziv-Zakai bound with arbitrary inputs. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2087–2092, 2023.

[9] Minoh Jeong, Alex Dytso, and Martina Cardone. A comprehensive study on Ziv-Zakai lower bounds on the MMSE. *arXiv preprint arXiv:2404.04366*, 2024.

[10] Minoh Jeong, Martina Cardone, and Alex Dytso. Demystifying the optimal performance of multi-class classification. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 31638–31664. Curran Associates, Inc., 2023.

[11] Minoh Jeong, Martina Cardone, and Alex Dytso. Data-driven estimation of the false positive rate of the Bayes binary classifier via soft labels. *arXiv preprint arXiv:2401.15500*, 2024.

[12] Mohammad Milanian, Minoh Jeong, and Martina Cardone. On the secrecy capacity of 1-2-1 atomic networks. *arXiv preprint arXiv:2405.05823*, 2024.

[13] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Characterization of the quality factor in spiral coil designs for high-frequency wireless power transfer systems using machine learning. In *2022 IEEE 23rd Workshop on Control and Modeling for Power Electronics (COMPEL)*, pages 1–8, 2022.

[14] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Optimization of spiral coil design for WPT systems using machine learning. In *2023 IEEE Applied Power Electronics Conference and Exposition (APEC)*, pages 822–828, 2023.

[15] Minki Kim, Minoh Jeong, Martina Cardone, and Jungwon Choi. Design of a spiral coil for high-frequency wireless power transfer systems using machine learning. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 5(1):193–202, 2024.

[16] Cynthia Dwork. Differential privacy: A survey of results. In Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, editors, *Theory and Applications of Models of Computation*, pages 1–19, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[17] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.

[18] Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 43–54, New York, NY, USA, 2016. Association for Computing Machinery.

[19] Shayle R Searle. Prediction, mixed models, and variance components. 1973.

[20] Stephen Portnoy. Maximizing the probability of correctly ordering random variables using linear predictors. *Journal of Multivariate Analysis*, 12(2):256–269, 1982.

[21] Kentaro Nomakuchi and Toshio Sakata. Characterizations of the forms of covariance matrix of an elliptically contoured distribution. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 50(2):205–210, 1988.

[22] Kentaro Nomakuchi and Toshio Sakata. Characterization of conditional covariance and unified theory in the problem of ordering random variables. *Annals of the Institute of Statistical Mathematics*, 40(1):93–99, 1988.

[23] O. Collier and A. S. Dalalyan. Minimax rates in permutation estimation for feature matching. *The Journal of Machine Learning Research*, 17(6):1 –31, January 2016.

[24] Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Transactions on Information Theory*, 64(5):3286–3300, 2018.

[25] Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Denoising linear models with permuted data. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 446–450, 2017.

[26] Philippe Rigollet and Jonathan Weed. Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Information and Inference: A Journal of the IMA*, 8(4):691–717, December 2019.

[27] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli. Unlabeled sensing with random linear measurements. *IEEE Transactions on Information Theory*, 64(5):3237–3253, May 2018.

[28] S. Haghighatshoar and G. Caire. Signal recovery from unlabeled samples. *IEEE Transactions on Signal Processing*, 66(5):1242–1257, March 2018.

[29] Ivan Dokmanić. Permutations unlabeled beyond sampling unknown. *IEEE Signal Processing Letters*, 26(6):823–827, April 2019.

[30] Manolis C. Tsakiris and Liangzu Peng. Homomorphic sensing. In *International Conference on Machine Learning*, pages 6335–6344. PMLR, 2019.

[31] Manolis C Tsakiris. Eigenspace conditions for homomorphic sensing. *arXiv:1812.07966*, April 2019.

[32] Hang Zhang, Martin Slawski, and Ping Li. Permutation recovery from multiple measurement vectors in unlabeled sensing. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 1857–1861, 2019.

[33] Hang Zhang, Martin Slawski, and Ping Li. The benefits of diversity: Permutation recovery in unlabeled sensing from multiple measurement vectors. *IEEE Transactions on Information Theory*, 68(4):2509–2529, 2022.

[34] Hang Zhang and Ping Li. Optimal estimator for unlabeled linear regression. In *International Conference on Machine Learning*, pages 11153–11162. PMLR, 2020.

[35] Liangzu Peng and Manolis C Tsakiris. Linear regression without correspondences via concave minimization. *IEEE Signal Processing Letters*, 27:1580–1584, 2020.

[36] Alex Dytso, Martina Cardone, Mishfad S. Veedu, and H. Vincent Poor. On estimation under noisy order statistics. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 36–40, 2019.

[37] Shang Shang, Tiance Wang, Paul Cuff, and Sanjeev Kulkarni. The application of differential privacy for rank aggregation: Privacy and accuracy. In *17th International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2014.

[38] Michael Hay, Liudmila Elagina, and Gerome Miklau. Differentially private rank aggregation. In *2017 SIAM International Conference on Data Mining*, pages 669–677. SIAM, 2017.

[39] Ziqi Yan, Gang Li, and Jiqiang Liu. Private rank aggregation under local differential privacy. *International Journal of Intelligent Systems*, 35(10):1492–1519, 2020, https://onlinelibrary.wiley.com/doi/pdf/10.1002/int.22261.

[40] Daniel Alabi, Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Private rank aggregation in central and local models. In *AAAI Conference on Artificial Intelligence*, volume 36, pages 5984–5991, 2022.

[41] T. Tony Cai and Rong Ma. Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. *IEEE Transactions on Information Theory*, pages 1–1, 2023.

[42] Manolis C. Tsakiris. Matrix recovery from permutations: An algebraic geometry approach. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2511–2516, 2023.

[43] Manolis C Tsakiris. Ladder matrix recovery from permutations. *arXiv preprint arXiv:2207.10864*, 2022.

[44] Hang Zhang and Ping Li. Optimal estimator for linear regression with shuffled labels. *arXiv preprint arXiv:2310.01326*, 2023.

[45] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[46] Martin Slawski, Emanuel Ben-David, and Ping Li. Two-stage approach to multivariate linear regression with sparsely mismatched data. *Journal of Machine Learning Research*, 21(204):1–42, 2020.

[47] Feiran Li, Kent Fujiwara, Fumio Okura, and Yasuyuki Matsushita. Generalized shuffled linear regression. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6474–6483, October 2021.

[48] Yunzhen Yao, Liangzu Peng, and Manolis C. Tsakiris. Unlabeled principal component analysis. *Advances in Neural Information Processing Systems*, 34, 2021.

[49] Zhiwei Tang, Tsung-Hui Chang, Xiaojing Ye, and Hongyuan Zha. Low-rank matrix recovery with unknown correspondence. In Robin J. Evans and Ilya Shpitser, editors, *Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 2111–2122. PMLR, 31 Jul–04 Aug 2023.

[50] Hang Zhang and Ping Li. Sparse recovery with shuffled labels: Statistical limits and practical estimators. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1760–1765. IEEE, 2021.

[51] Rahul Mazumder and Haoyue Wang. Linear regression with mismatched data: A provably optimal local search algorithm. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 443–457. Springer, 2021.

[52] Martin Slawski, Guoqing Diao, and Emanuel Ben-David. A pseudo-likelihood approach to linear regression with partially shuffled data. *Journal of Computational and Graphical Statistics*, 30(4):991–1003, 2021.

[53] Herbert Aron David and Haikady Navada Nagaraja. *Order Statistics*. Wiley Online Library, 2004.

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

[55] Steven M Kay. *Fundamentals of Statistical Signal Processing, vol. 1: Estimation Theory*. Prentice Hall PTR, 1993.

[56] Peter M Gruber. *Convex and Discrete Geometry*, volume 336. Springer Science & Business Media, 2007.

[57] J Bourgain, J Lindenstrauss, and V Milman. Estimates related to Steiner symmetrizations. In *Geometric Aspects of Functional Analysis*, pages 264–273. Springer, 1989.

[58] Daniel A Klain. Steiner symmetrization using a finite set of directions. *Advances in Applied Mathematics*, 48(2):340–353, 2012.

[59] William Clement Karl, George C Verghese, and Alan S Willsky. Reconstructing ellipsoids from projections. *CVGIP: Graphical Models and Image Processing*, 56(2):124–139, March 1994.

[60] R. Pyke. Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(3):395–449, 1965.

[61] James Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems*, 25, 2012.

[62] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford university press, 2013.

[63] N. Balakrishnan, V.B. Nevzorov, and A. Stepanov. On normal spacings. *Statistics & Probability Letters*, 193:109713, 2023.

[64] Ravi Kumar and Sergei Vassilvitskii. Generalized distances between rankings. In *19th International Conference on World Wide Web*, pages 571–580, 2010.

[65] Moshe Shaked and YL Tong. Stochastic ordering of spacings from dependent random variables. *Lecture Notes-Monograph Series*, pages 141–149, 1984.

[66] Moshe Shaked and J George Shanthikumar. *Stochastic Orders*. Springer, 2007.

[67] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

[68] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006.

[69] Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

[70] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454–3469, 2020.

[71] Brendan Avent, Javier González, Tom Diethe, Andrei Paleyes, and Borja Balle. Automatic discovery of privacy–utility Pareto fronts. *Proceedings on Privacy Enhancing Technologies*, 4:5–23, 2020.

[72] Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, 2015.

[73] Jordi Soria-Comas and Josep Domingo-Ferrer. Optimal data-independent noise for differential privacy. *Information Sciences*, 250:200–214, 2013.

[74] Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Optimal noise-adding mechanism in additive differential privacy. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 11–20. PMLR, 2019.

[75] Saralees Nadarajah. A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694, 2005.

[76] Alex Dytso, Ronit Bustin, H Vincent Poor, and Shlomo Shamai. Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*, 5(1):1–40, 2018.

[77] Fang Liu. Generalized Gaussian mechanism for differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):747–756, 2018.

[78] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[79] Manuel Gil, Fady Alajaji, and Tamas Linder. Rényi divergence measures for commonly used univariate continuous distributions. *Information Sciences*, 249:124–131, 2013.

[80] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web*, pages 613–622, 2001.

[81] Stefan Chanas and Prezemysław Kobylański. A new heuristic algorithm solving the linear ordering problem. *Computational Optimization and Applications*, 6(2):191–205, 1996.

[82] Jacob P Baskin and Shriram Krishnamurthi. Preference aggregation in group recommender systems for committee decision-making. In *Proceedings of the third ACM Conference on Recommender Systems*, pages 337–340, 2009.

[83] Serdar Özyurt and Oğuz Kucur. Power permutation modulation in multiple-input multiple-output systems. *Transactions on Emerging Telecommunications Technologies*, page e4408, 2021.

[84] Fuzhen Zhang. *The Schur Complement and Its Applications*, volume 4. Springer Science & Business Media, 2006.

[85] Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295, 1983.

[86] Jan Holstermann. A generalization of the rearrangement inequality. *Mathematical Reflections*, 5(4), 2017.

[87] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[88] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1964.

[89] Henryk Minc and Leroy Sathre. Some inequalities involving (r!) 1/r. *Proceedings of the Edinburgh Mathematical Society*, 14(1):41–46, 1964.

Appendix A

Differed Proofs in Chapter 3

A.1 Proof of Proposition 3.3.1

We start by noting that any $\pi_1 \in \mathcal{P}$ has its own unique $\pi_2 \in \mathcal{P}$ such that $\mathcal{H}_{\pi_1} = -\mathcal{H}_{\pi_2}$ Then, for any observation $\mathbf{y}$, we have that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\pi_1}) &= \int_{\mathbf{x} \in \mathcal{H}_{\pi_1}} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\overset{(a)}{=} \int_{\mathbf{z} \in -\mathcal{H}_{\pi_1}} f_{\mathbf{N}}(\mathbf{y} + \mathbf{z}) f_{\mathbf{X}}(\mathbf{z}) \, \mathrm{d}\mathbf{z} \\ &\overset{(b)}{=} \int_{\mathbf{z} \in \mathcal{H}_{\pi_2}} f_{\mathbf{N}}(-\mathbf{y} - \mathbf{z}) f_{\mathbf{X}}(\mathbf{z}) \, \mathrm{d}\mathbf{z} \\ &= f_{\mathbf{Y}}(-\mathbf{y}, \mathcal{H}_{\pi_2}), \end{aligned} \tag{A.1}$$

where the labeled equalities follow from: (a) change of variable $\mathbf{z} = -\mathbf{x}$; and (b) the fact that $\mathcal{H}_{\pi_1} = -\mathcal{H}_{\pi_2}$ and $f_{\mathbf{N}}(\mathbf{n}) = f_{\mathbf{N}}(-\mathbf{n})$.

From the relation in (A.1), it therefore follows that we can map $f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\pi_1})$ to $f_{\mathbf{Y}}(-\mathbf{y}, \mathcal{H}_{\pi_2})$ for all (π_1, π_2) index pairs where $\pi_1 \in \mathcal{P}$ and $\pi_2 \in \mathcal{P}$ such that $\mathcal{H}_{\pi_1} = -\mathcal{H}_{\pi_2}$. Assume now that $\mathbf{y} \in \mathcal{R}_{\pi_1, K_{\mathbf{N}}}$, which from (3.4) implies that $f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\pi_1})$ is the maximum among all $f_{\mathbf{Y}}(\mathbf{y}, \mathcal{H}_{\tau}),\ \tau \in \mathcal{P}$. From (A.1) we then have that, among all $f_{\mathbf{Y}}(-\mathbf{y}, \mathcal{H}_{\tau}),\ \tau \in \mathcal{P}$, the maximum joint density for $-\mathbf{y}$ is $f_{\mathbf{Y}}(-\mathbf{y}, \mathcal{H}_{\pi_2})$ where π_2 is such that $\mathcal{H}_{\pi_2} = -\mathcal{H}_{\pi_1}$. This, from (3.4), implies that

$$-\mathbf{y} \in \mathcal{R}_{\pi_2, K_{\mathbf{N}}}. \tag{A.2}$$

This concludes the proof of Proposition 3.3.1.

A.2 Proof of Proposition 3.4.5

Let $\lambda_i,\ i \in [1:n]$ be the eigenvalues of $K_{\mathbf{N}}$ and $\tilde{\lambda}_i,\ i \in [1:n]$ be the eigenvalues of $(K_{\mathbf{N}}^{-1} + I_n)^{-1}$ in (3.6). Then, for all $i \in [1:n]$, the relationship between λ_i and $\tilde{\lambda}_i$ is such that

$$\lambda_i = \frac{\tilde{\lambda}_i}{1 - \tilde{\lambda}_i}. \tag{A.3}$$

For the case when $K_{\mathbf{N}}$ has n equal eigenvalues (i.e., either $K_{\mathbf{N}}$ is a diagonal matrix with equal elements on the diagonal, or we take $v = 0$ and $\gamma = a$ in (3.6)), it is not difficult to verify that the eigenvalues and eigenvectors are of the form in (3.9).

We hence focus on the case when $K_{\mathbf{N}}$ has at least two distinct eigenvalues (i.e., when either $v = 0, \gamma \neq a$, or when $v \neq 0$ in (3.6)). Since $(K_{\mathbf{N}}^{-1} + I_n)^{-1}$ in (3.6) consists of an orthonormal matrix $Q \in \mathcal{Q}$ and a block diagonal matrix, its eigenvalues can be found as the solution of

$$\tilde{\lambda}_i = \gamma,\ i \in [1:n-2], \tag{A.4a}$$

$$a = \tilde{\lambda}_{n-1} + \tilde{\lambda}_n - \gamma, \tag{A.4b}$$

$$v^2 = (\tilde{\lambda}_{n-1} - \gamma)(\gamma - \tilde{\lambda}_n), \tag{A.4c}$$

where the second expression is due to the fact that $\gamma + a = \tilde{\lambda}_{n-1} + \tilde{\lambda}_n$ and the last expression follows by computing the determinant of S in (3.6) with (A.4b). By solving the above set of linear equations and by using (A.3) we obtain the eigenvalues in (3.10a) – see also Appendix A.10.

We now use the eigenvalues in (3.10a) to find the eigenvectors $\boldsymbol{\nu}_i$ of $K_{\mathbf{N}}$. We start by noting that $\boldsymbol{\nu}_i$'s are equal to the eigenvectors of $(K_{\mathbf{N}}^{-1} + I_n)^{-1}$. Since the matrix in (3.6) has one isotropic matrix, we can easily find the first $n-2$ eigenvectors of $(K_{\mathbf{N}}^{-1} + I_n)^{-1}$ (i.e., those associated to the eigenvalue γ) as,

$$\boldsymbol{\nu}_i = \mathbf{q}_i,\ i \in [1:n-2], \tag{A.5}$$

where $\mathbf{q}_i$ is i-th column of $Q \in \mathcal{Q}$. For $\boldsymbol{\nu}_i,\ i \in [n-1:n]$, by using the eigendecomposition of S and the fact that $(K_{\mathbf{N}}^{-1} + I_n)^{-1} = QV\tilde{\Lambda}V^TQ^T$ with $\tilde{\Lambda}$ being a diagonal matrix and V being

an orthonormal matrix, we obtain the following two equations,

$$\boldsymbol{\nu}_i = (a - \tilde{\lambda}_i)\mathbf{q}_{n-1} + v\mathbf{q}_n, i \in [n-1:n], \tag{A.6}$$

$$\boldsymbol{\nu}_i = v\mathbf{q}_{n-1} + (\gamma - \tilde{\lambda}_i)\mathbf{q}_n, i \in [n-1:n]. \tag{A.7}$$

By combining (A.6) and (A.7), and by using (A.3) we obtain the eigenvectors in (3.10b). This concludes the proof of Proposition 3.4.5.

A.3 Proof of Proposition 3.4.9

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 3.4.1, the probability of correctness can be written as

$$\begin{aligned}
P_c &= \sum_{\pi\in\mathcal{P}} \Pr\left((\mathbf{X},\mathbf{Y})^T \in \mathcal{H}_\pi \times \mathcal{R}_{\pi,K_\mathbf{N}}\right) \\
&\overset{(a)}{=} \sum_{\pi\in\mathcal{P}} \Pr\left((\mathbf{X},\mathbf{Y})^T \in \mathcal{H}_\pi \times (K_\mathbf{N}+I_n)\mathcal{H}_\pi\right) \\
&\overset{(b)}{=} \sum_{\pi\in\mathcal{P}} \Pr\left(\left(\mathbf{X},\mathbf{X}+K_\mathbf{N}^{\frac{1}{2}}\mathbf{Z}\right)^T \in \mathcal{H}_\pi \times (K_\mathbf{N}+I_n)\mathcal{H}_\pi\right) \\
&\overset{(c)}{=} \sum_{\pi\in\mathcal{P}} \Pr\left(A(\mathbf{X},\mathbf{Z})^T \in \mathcal{H}_\pi \times (K_\mathbf{N}+I_n)\mathcal{H}_\pi\right) \\
&\overset{(d)}{=} \sum_{\pi\in\mathcal{P}} \Pr\left((\mathbf{X},\mathbf{Z})^T \in A^{-1}\mathcal{C}_{\mathcal{H}_\pi}\right) \\
&\overset{(e)}{=} n!\Pr\left((\mathbf{X},\mathbf{Z})^T \in A^{-1}\mathcal{C}_{\mathcal{H}_\pi}\right),
\end{aligned} \tag{A.8}$$

where the labeled equalities follow from: (a) using the optimal decision regions in Theorem 3.4.1; (b) letting $\mathbf{Z}$ be a standard normal random vector, i.e., $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$; (c) defining $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & K_\mathbf{N}^{\frac{1}{2}} \end{bmatrix}$; (d) letting $\mathcal{C}_{\mathcal{H}_\pi} = \mathcal{H}_\pi \times (K_\mathbf{N}+I_n)\mathcal{H}_\pi$; and (e) using the symmetry of $(\mathbf{X},\mathbf{Z})$.

We observe that the shape of the region $\mathcal{H}_\pi$ is an n-dimensional cone (see Fig. 2.2 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 $(\mathbf{X}, \mathbf{Z})^T$ to fall within a cone. Using the symmetry of the Gaussian distribution, the probability of a pair $(\mathbf{X}, \mathbf{Z})^T$ 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

$$\begin{aligned}\Pr\left((\mathbf{X}, \mathbf{Z})^T \in A^{-1}\mathcal{C}_{\mathcal{H}_\pi}\right) &= \frac{\mathrm{Vol}^{2n}\left(A^{-1}\mathcal{C}_{\mathcal{H}_\pi} \cap \mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}{\mathrm{Vol}^{2n}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)} \\ &= \frac{|\det(A^{-1})|\mathrm{Vol}^{2n}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}{\mathrm{Vol}^{2n}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}, \end{aligned} \tag{A.9}$$

where in the last equality we have used the fact that $\mathrm{Vol}^k\left(A\mathcal{S}\right) = |\det(A)|\mathrm{Vol}^k\left(\mathcal{S}\right)$ for any invertible matrix A and any set $\mathcal{S}$. By combining (A.8) and (A.9) we arrive at

$$P_c = n!\frac{|\det(A^{-1})|\mathrm{Vol}^{2n}\left(\mathcal{C}_{\mathcal{H}_\pi} \cap A\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}{\mathrm{Vol}^{2n}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}. \tag{A.10}$$

The proof of Proposition 3.4.9 is concluded by noting that A is a block matrix and hence $|\det(A)| = \det\left(K_{\mathbf{N}}^{\frac{1}{2}}\right)$, and by using the fact that $P_e = 1 - P_c$.

A.4 Proof of Lemma 3.5.4

We start by observing that, since $K_{\mathbf{U}}$ is positive definite, we have that

$$\begin{aligned}\Pr(\mathbf{U} \in \mathcal{H}_\pi) &= \Pr\left(K_{\mathbf{U}}^{\frac{1}{2}}\mathbf{Z} \in \mathcal{H}_\pi\right) \\ &= \Pr\left(\mathbf{Z} \in K_{\mathbf{U}}^{-\frac{1}{2}}\mathcal{H}_\pi\right) \\ &= \frac{\mathrm{Vol}^n\left(K_{\mathbf{U}}^{-\frac{1}{2}}\mathcal{H}_\pi \cap \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\mathrm{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}, \end{aligned} \tag{A.11}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$, and where the last equality follows by representing the probability in terms of a ratio of two volumes.

We then obtain

$$\mathrm{Vol}^n\left(K_{\mathbf{U}}^{-\frac{1}{2}}\mathcal{H}_\pi \cap \mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right) = \left|\det\left(K_{\mathbf{U}}^{-\frac{1}{2}}\right)\right| \mathrm{Vol}^n\left(\mathcal{H}_\pi \cap K_{\mathbf{U}}^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right), \tag{A.12}$$

where the equality follows from the fact that, for an $n \times n$ invertible matrix A and a set $\mathcal{S} \subseteq \mathbb{R}^n$, we have that $\mathrm{Vol}^n(A\mathcal{S}) = |\det(A)|\mathrm{Vol}^n(\mathcal{S})$. Finally, by substituting (A.12) into (A.11) we obtain

$$\Pr(\mathbf{U} \in \mathcal{H}_\pi) = \frac{\left|\det\left(K_{\mathbf{U}}^{-\frac{1}{2}}\right)\right| \mathrm{Vol}^n\left(\mathcal{H}_\pi \cap K_{\mathbf{U}}^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\mathrm{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}. \tag{A.13}$$

This concludes the proof of Lemma 3.5.4.

A.5 Proof of Lemma 3.5.5

We start by observing that, from the definition of the optimal decision regions in (3.4), we have that $\mathbf{0}_n \in \bigcap_{\pi\in\mathcal{P}} \mathcal{R}_{\pi,K_{\mathbf{N}}}$ if and only if

$$f_{\mathbf{Y}}(\mathbf{0}_n, \mathcal{H}_\pi) = d, \ \forall \pi \in \mathcal{P}, \tag{A.14}$$

for some constant $d > 0$. Note that this implies that

$$\Pr(\mathbf{X} \in \mathcal{H}_\pi | \mathbf{Y} = \mathbf{0}_n) = d', \ \forall \pi \in \mathcal{P}, \tag{A.15}$$

where $d' = d/f_{\mathbf{Y}}(\mathbf{0}_n)$. Furthermore, recall that $\mathbf{X}|\mathbf{Y} = \mathbf{y}$ is Gaussian (see Remark 3.4.2) and for any $\mathbf{y} \in \mathbb{R}^n$

$$\Pr\left(\mathbf{X} \in \mathcal{H}_\pi | \mathbf{Y} = \mathbf{y}\right) = \Pr\left(\left(I_n + K_{\mathbf{N}}^{-1}\right)^{-1}\mathbf{y} + \left(I_n + K_{\mathbf{N}}^{-1}\right)^{-\frac{1}{2}}\mathbf{Z} \in \mathcal{H}_\pi\right), \forall \pi \in \mathcal{P}, \tag{A.16}$$

where $\mathbf{Z}$ is a standard Gaussian random vector. Now, by evaluating (A.16) and combining it with Lemma 3.5.4, we have that

$$\Pr(\mathbf{X} \in \mathcal{H}_\pi | \mathbf{Y} = \mathbf{0}_n) = \frac{\left|\det(\mathring{K}^{-\frac{1}{2}})\right| \mathrm{Vol}^n\left(\mathcal{H}_\pi \cap \mathring{K}^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\mathrm{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}, \tag{A.17}$$

where $\mathring{K} = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$. Finally, the sufficient and necessary condition in (A.15) together with (A.17), imply that

$$\frac{\left|\det(\mathring{K}^{-\frac{1}{2}})\right| \operatorname{Vol}^n\left(\mathcal{H}_\pi \cap \mathring{K}^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\operatorname{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)} = d', \forall \pi \in \mathcal{P}, \tag{A.18}$$

which, after rescaling and substituting $\mathring{K} = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$, reduces to (3.26) where

$$\eta = d' \frac{\operatorname{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\left|\det(\mathring{K}^{-\frac{1}{2}})\right|}.$$

This concludes the proof of Lemma 3.5.5.

A.6 Proof of Lemma 3.5.7

We start by noting that the proof of Lemma 3.5.7 for the case $n = 2$ is immediate, and hence we next focus on the case $n > 2$. In particular, our proof will leverage an auxiliary result presented in the next lemma, the proof of which can be found in Appendix A.9.

Lemma A.6.1. *Let $\mathcal{E}^n$ be an n-dimensional ellipsoid centered at the origin with unitary axes $\{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_n\}$ and corresponding radii equal to $\{r_1, r_2, \ldots, r_n\}$. Moreover, for $r \in \mathbb{R}$, define the following hyperplane and $n-1$ dimensional ellipsoid:*

$$\mathcal{W}(r) = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\nu}_n^T \mathbf{x} = r\}, \tag{A.19}$$

$$\mathcal{E}_{\mathcal{W}(r)}^{n-1} = \mathcal{E}^n \cap \mathcal{W}(r). \tag{A.20}$$

If $\boldsymbol{\nu}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n$, then for every $\pi \in \mathcal{P}$

$$\operatorname{Vol}^n\left(\mathcal{H}_\pi \cap \mathcal{E}^n\right) = \operatorname{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}_{\mathcal{W}(0)}^{n-1}\right) c(r_n), \tag{A.21}$$

where $c(r_n)$ is a constant that only depends on r_n.

By leveraging Lemma A.6.1, for a constant $\eta > 0$, we have that

$$\operatorname{Vol}^n\left(\mathcal{H}_\pi \cap \mathcal{E}^n\right) = \eta, \ \forall \pi \in \mathcal{P}, \tag{A.22}$$

if and only if

$$\mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) = \tilde{\eta},\ \forall \pi \in \mathcal{P}, \tag{A.23}$$

where $\mathcal{E}^{n-1}_{\mathcal{W}(0)} = \mathcal{E}^n \cap \mathcal{W}(0)$, and where $\tilde{\eta}$ is some other constant. Therefore, if (A.22) holds then so does (A.23) and vice versa. Consequently, to prove Lemma 3.5.7, we need to show that (A.23) holds if and only if $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ is an $(n-1)$-dimensional ball. Remember that $\mathcal{E}^{n-1}_{\mathcal{W}(0)} \subset \mathcal{W}(0)$ has unitary axes $\{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{n-1}\}$ with corresponding radii equal to $\{r_1, r_2, \ldots, r_{n-1}\}$.

First, suppose that $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ is an $(n-1)$-dimensional ball. Then, from the symmetry of $\mathcal{H}_\pi$'s, it readily follows that (A.23) holds (and hence (A.22) holds). Hence, the fact that $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ is an $(n-1)$-dimensional ball is a sufficient condition for (A.22) to hold. We now show that it is also necessary. In particular, our proof follows by using a contradiction argument where we assume that $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ is not an $(n-1)$-dimensional ball.

Assume that $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ has at least one radius that is different from the others. Without loss of generality, let $r_1 = \max_{i\in[1:n-1]}\{r_i\}$ and $r_2 = \min_{i\in[1:n-1]}\{r_i\}$. Note that $r_1\boldsymbol{\nu}_1 \in \mathcal{E}^{n-1}_{\mathcal{W}(0)}$ and $r_2\boldsymbol{\nu}_2 \in \mathcal{E}^{n-1}_{\mathcal{W}(0)}$. Assume that $r_1\boldsymbol{\nu}_1 \in \mathcal{H}_\alpha$ and $r_2\boldsymbol{\nu}_2 \in \mathcal{H}_\beta$, for some $\alpha, \beta \in \mathcal{P}$. Note that $\alpha \neq \beta$, i.e., when $n > 2$, there is no possibility for any of the $\mathcal{H}_\pi$'s to contain more than one axis of $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$. Next, observe that $\mathcal{H}_\alpha \cap \mathcal{W}(0)$ and $\mathcal{H}_\beta \cap \mathcal{W}(0)$ have equal $(n-1)$-dimensional cone shapes (i.e., the angular measures of the two cones are the same) in the subspace $\mathcal{W}(0)$. We let $\mathcal{B}^{n-1}_{\mathcal{W}}(\mathbf{0}_n, r) = \mathcal{B}^n(\mathbf{0}_n, r) \cap \mathcal{W}(0)$ be the $(n-1)$-dimensional ball of radius r. Because of the assumption of $r_1 \neq r_2$, there exists some value $\tilde{r}$, such that $r_1 > \tilde{r} > r_2$ and

$$\begin{aligned}
\mathrm{Vol}^{n-1}\left(\mathcal{H}_\alpha \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) &\overset{(a)}{=} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\alpha \cap \mathcal{W}(0) \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) \\
&\overset{(b)}{>} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\alpha \cap \mathcal{W}(0) \cap \mathcal{B}^{n-1}_{\mathcal{W}}(\mathbf{0}_n, \tilde{r})\right) \\
&\overset{(c)}{=} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\beta \cap \mathcal{W}(0) \cap \mathcal{B}^{n-1}_{\mathcal{W}}(\mathbf{0}_n, \tilde{r})\right) \\
&\overset{(d)}{>} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\beta \cap \mathcal{W}(0) \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) \\
&= \mathrm{Vol}^{n-1}\left(\mathcal{H}_\beta \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right),
\end{aligned} \tag{A.24}$$

where the labeled (in)equalities follow from: (a) the fact that $\mathcal{E}^{n-1}_{\mathcal{W}(0)} \subset \mathcal{W}(0)$; (b) the assumption that the cone $\mathcal{H}_\alpha \cap \mathcal{W}(0)$ contains the largest axis of the ellipsoid (i.e., $r_1\boldsymbol{\nu}_1 \in \mathcal{H}_\alpha$) and the assumption $\tilde{r} < r_1$; (c) using the fact that $\mathcal{H}_\beta, \mathcal{H}_\alpha, \mathcal{B}^{n-1}_{\mathcal{W}}(\mathbf{0}_n, \tilde{r})$ and $\mathcal{W}(0)$ are permutation

invariant; and (d) the assumption that the cone $\mathcal{H}_\beta \cap \mathcal{W}(0)$ contains the smallest axis of the ellipsoid (i.e., $r_2\boldsymbol{\nu}_2 \in \mathcal{H}_\beta$) and the assumption $\tilde{r} > r_2$.

This shows that, if $r_1 \neq r_2$, then (A.23) (and hence (A.22)) can not hold. Therefore, for (A.23) (and hence (A.22)) to hold, $\mathcal{E}^{n-1}_{\mathcal{W}(0)}$ must be an $(n-1)$-dimensional ball, i.e., the radii $\{r_1, \ldots, r_{n-1}\}$ of $\mathcal{E}^n$ must be all equal to each other. This concludes the proof of Lemma 3.5.7.

A.7 Sufficient and Necessary Conditions for Lemma 3.5.8

We start by noting that, by substituting $B = \gamma I_{n-1}$ inside (3.36), we obtain

$$I_{n-1}\gamma = C^T \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} C. \tag{A.25}$$

Moreover, we also note that

$$\begin{aligned} CC^T &\overset{(a)}{=} \begin{bmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_{n-1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_{n-1}^T \end{bmatrix} \\ &\overset{(b)}{=} \begin{bmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_n \end{bmatrix} \left(I_n - \begin{bmatrix} 0_{(n-1)\times(n-1)} & \mathbf{0}_{n-1} \\ \mathbf{0}_{n-1}^T & 1 \end{bmatrix} \right) \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_n^T \end{bmatrix} \\ &\overset{(c)}{=} I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T = I_n - \frac{1}{n}1_{n\times n}, \end{aligned} \tag{A.26}$$

where the labeled equalities follow from: (a) letting $\mathbf{c}_i, i \in [1 : n-1]$ be the i-th column of C; (b) letting $\mathbf{c}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n$; and (c) noting that $\mathbf{c}_n$ is a unit vector that belongs to $\mathcal{L}_{\mathcal{H}}$ in (3.27) and hence, it is perpendicular to $\mathcal{W}$ and to its orthonormal basis formed by the $n-1$ columns of C.

Now recall that the set $\mathcal{Q}$ is the set of $n \times n$ real-valued orthonormal matrices with the n-th column equal to $\frac{1}{\sqrt{n}}\mathbf{1}_n$. Moreover, note that since the matrix C in (A.25) is *any* orthonormal matrix the columns of which form a basis of the hyperplane $\mathcal{W}$, then the matrix $Q \in \mathcal{Q}$ can be chosen so as to have C to populate its first $n-1$ columns. In other words, we can always find a pair (Q, C) with $Q \in \mathcal{Q}$ such that

$$Q = \begin{bmatrix} C & \frac{1}{\sqrt{n}}\mathbf{1}_n \end{bmatrix}. \tag{A.27}$$

Without loss of generality, we assume the structure in (A.27) for Q, and we let

$$\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} = QAQ^T. \tag{A.28}$$

Note that the matrix A in (A.28) is symmetric. This follows from the fact that the left-hand side of (A.28) is positive definite, and hence symmetric. This implies that $QAQ^T = (QAQ^T)^T$, which leads to $A = A^T$. Then, we obtain

$$C^T \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} C = C^T QAQ^T C = \begin{bmatrix} I_{n-1} & \mathbf{0}_{n-1} \end{bmatrix} A \begin{bmatrix} I_{n-1} \\ \mathbf{0}_{n-1}^T \end{bmatrix},$$

and hence from (A.25), we need

$$\gamma I_{n-1} = \begin{bmatrix} I_{n-1} & \mathbf{0}_{n-1} \end{bmatrix} A \begin{bmatrix} I_{n-1} \\ \mathbf{0}_{n-1}^T \end{bmatrix},$$

which implies that A has to have the form as

$$A = \begin{bmatrix} \gamma I_{n-1} & \mathbf{v} \\ \mathbf{v}^T & a \end{bmatrix},$$

for some constant a and column vector $\mathbf{v}$ of dimension $n-1$. By substituting this back into (A.28), we obtain

$$\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} = Q \begin{bmatrix} \gamma I_{n-1} & \mathbf{v} \\ \mathbf{v}^T & a \end{bmatrix} Q^T, \tag{A.29}$$

where $Q \in \mathcal{Q}$. Moreover, since we can arbitrarily choose the first $n-1$ columns of $Q \in \mathcal{Q}$, the expression in (A.29) can be further simplified as

$$\begin{aligned} \left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} &= Q \begin{bmatrix} \gamma I_{n-1} & \mathbf{0}_{n-1} \\ \mathbf{0}_{n-1}^T & a \end{bmatrix} Q^T + Q \begin{bmatrix} 0_{(n-1)\times(n-1)} & \mathbf{v} \\ \mathbf{v}^T & 0 \end{bmatrix} Q^T \\ &\overset{(a)}{=} \tilde{Q} \begin{bmatrix} \gamma I_{n-1} & \mathbf{0}_{n-1} \\ \mathbf{0}_{n-1}^T & a \end{bmatrix} \tilde{Q}^T + \tilde{Q} \begin{bmatrix} 0_{(n-2)\times(n-2)} & 0_{(n-2)\times 2} \\ 0_{2\times(n-2)} & D \end{bmatrix} \tilde{Q}^T \\ &= \tilde{Q} \begin{bmatrix} \gamma I_{n-2} & 0_{(n-2)\times 2} \\ 0_{2\times(n-2)} & S \end{bmatrix} \tilde{Q}^T, \end{aligned} \tag{A.30}$$

where $S = \left[\begin{smallmatrix} \gamma & v \\ v & a \end{smallmatrix}\right]$ with $v \in \mathbb{R}$, and where the equality in (a) follows since, for $Q \in \mathcal{Q}$ we have that

$$
\begin{aligned}
Q \begin{bmatrix} 0_{(n-1)\times(n-1)} & \mathbf{v} \\ \mathbf{v}^T & 0 \end{bmatrix} Q^T &= \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_{n-1} & \frac{\mathbf{1}_n}{\sqrt{n}} \end{bmatrix} \begin{bmatrix} 0_{(n-1)\times(n-1)} & \mathbf{v} \\ \mathbf{v}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^T \\ \vdots \\ \mathbf{c}_{n-1}^T \\ \frac{\mathbf{1}_n^T}{\sqrt{n}} \end{bmatrix} \\
&\overset{(\mathrm{a1})}{=} \frac{\mathbf{1}_n}{\sqrt{n}} \left(\sum_{i=1}^{n-1} v_i \mathbf{c}_i^T \right) + \left(\sum_{i=1}^{n-1} v_i \mathbf{c}_i \right) \frac{\mathbf{1}_n^T}{\sqrt{n}} \\
&\overset{(\mathrm{a2})}{=} \frac{\mathbf{1}_n}{\sqrt{n}} v \tilde{\mathbf{c}}_{n-1}^T + v \tilde{\mathbf{c}}_{n-1} \frac{\mathbf{1}_n^T}{\sqrt{n}} \\
&\overset{(\mathrm{a3})}{=} \begin{bmatrix} \tilde{\mathbf{c}}_1 & \cdots & \tilde{\mathbf{c}}_{n-1} & \frac{\mathbf{1}_n}{\sqrt{n}} \end{bmatrix} \begin{bmatrix} 0_{(n-2)\times(n-2)} & 0_{(n-2)\times n} \\ 0_{2\times(n-2)} & D \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{c}}_1^T \\ \vdots \\ \tilde{\mathbf{c}}_{n-1}^T \\ \frac{\mathbf{1}_n^T}{\sqrt{n}} \end{bmatrix} \\
&\overset{(\mathrm{a4})}{=} \tilde{Q} \begin{bmatrix} 0_{(n-2)\times(n-2)} & 0_{(n-2)\times 2} \\ 0_{2\times(n-2)} & D \end{bmatrix} \tilde{Q}^T,
\end{aligned} \tag{A.31}
$$

where the labeled equalities follow from: $(\mathrm{a1})$ letting $v_i, i \in [1 : n-1]$ be the i-th element of $\mathbf{v}$; $(\mathrm{a2})$ noting that we can express $\sum_{i=1}^{n-1} v_i \mathbf{c}_i = v \tilde{\mathbf{c}}_{n-1}$ where v is a scalar and $\tilde{\mathbf{c}}_{n-1} \in \mathcal{W}$ is a unit vector orthogonal to $\frac{1}{\sqrt{n}}\mathbf{1}_n$; $(\mathrm{a3})$ the fact that $\tilde{\mathbf{c}}_1, \ldots, \tilde{\mathbf{c}}_{n-1}$ is an orthonormal basis of the hyperplane $\mathcal{W}$ and using matrix form representation; and $(\mathrm{a4})$ the fact that $\begin{bmatrix} \tilde{\mathbf{c}}_1 & \cdots & \tilde{\mathbf{c}}_{n-1} & \frac{1}{\sqrt{n}}\mathbf{1}_n \end{bmatrix} \in \mathcal{Q}$, and letting $D = \left[\begin{smallmatrix} 0 & v \\ v & 0 \end{smallmatrix}\right]$.

Thus, from (A.30) we have that

$$
\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} = \tilde{Q} \underbrace{\begin{bmatrix} \gamma I_{n-2} & 0_{n-2\times 2} \\ 0_{2\times n-2} & S \end{bmatrix}}_{B} \tilde{Q}^T. \tag{A.32}
$$

Since $\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1}$ is a positive definite matrix, we need to ensure that the Schur complement [84] of the block γI_{n-2} of the matrix B, denoted as $B/\gamma I_{n-2}$, is positive definite. Formally,

$$
B/\gamma I_{n-2} = S \text{ is positive definite } \implies a\gamma > v^2. \tag{A.33}
$$

We also need to find the conditions that ensure that $K_{\mathbf{N}}$ is positive definite. Towards this end, we perform the eigendecomposition of the matrix B, i.e., $B = V\Lambda V^T$, and rewrite (A.32) as

$$\left(K_{\mathbf{N}}^{-1} + I_n\right)^{-1} = QV\Lambda V^T Q^T, \tag{A.34}$$

where we highlight that the matrix QV is orthonormal. Thus,

$$\begin{aligned} &K_{\mathbf{N}}^{-1} + I_n = (QV\Lambda V^T Q^T)^{-1} = QV\Lambda^{-1}V^T Q^T \\ &\implies K_{\mathbf{N}}^{-1} = QV\Lambda^{-1}V^T Q^T - I_n = QV(\Lambda^{-1} - I_n)V^T Q^T \\ &\implies K_{\mathbf{N}} = QV(\Lambda^{-1} - I_n)^{-1}V^T Q^T. \end{aligned} \tag{A.35}$$

In order to ensure that $K_{\mathbf{N}}$ is positive definite, we compute its eigenvalues, which are given by the diagonal elements of the diagonal matrix $(\Lambda^{-1} - I_n)^{-1}$ and we find the conditions under which these are positive. Note that these correspond to the conditions for which Λ (i.e., the diagonal matrix with the eigenvalues of B) has diagonal elements strictly smaller than one. The eigenvalues of B are computed in Appendix A.10, where we have shown that B has $n-2$ eigenvalues equal to γ and the remaining two eigenvalues equal to

$$\lambda = \frac{a + \gamma \pm \sqrt{(a-\gamma)^2 + 4v^2}}{2}.$$

These eigenvalues must be strictly smaller than one, i.e.,

$$\gamma < 1, \tag{A.36a}$$

and

$$\begin{aligned} \tfrac{a+\gamma\pm\sqrt{(a-\gamma)^2+4v^2}}{2} < 1 \implies &\sqrt{(a-\gamma)^2 + 4v^2} < 2 - a - \gamma \\ \implies &v^2 < (1-a)(1-\gamma). \end{aligned} \tag{A.36b}$$

Note also that since $v^2 \geq 0$, we need $a < 1$. The expression in (A.32) together with the conditions in (A.33) and (A.36) conclude the proof of Lemma 3.5.8.

A.8 Proof of Lemma 3.5.9

From the result in Lemma 3.5.4, we have that for all $\pi \in \mathcal{P}$,

$$\Pr\left(\tilde{\mathbf{Y}}_0 \in \mathcal{H}_\pi\right) = \frac{\left|\det\left(\tilde{K}^{-\frac{1}{2}}\right)\right| \operatorname{Vol}^n\left(\mathcal{H}_\pi \cap \tilde{K}^{\frac{1}{2}}\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)}{\operatorname{Vol}^n\left(\mathcal{B}^n\left(\mathbf{0}_n, 1\right)\right)},$$

which together with Lemma 3.5.5 lead to the proof of (3.37). Now, note that (3.37) implies that for all $\pi \in \mathcal{P}$,

$$\beta = \Pr\left(\tilde{\mathbf{Y}}_0 \in \mathcal{H}_\pi\right) = \Pr\left(\mathbf{Z} \in \tilde{K}^{-\frac{1}{2}}\mathcal{H}_\pi\right) = \Pr\left(\mathbf{Z} \in \mathcal{C}_\pi\right), \tag{A.37}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ and $\mathcal{C}_\pi = \tilde{K}^{-\frac{1}{2}}\mathcal{H}_\pi$, $\forall \pi \in \mathcal{P}$. This further implies that $\mathcal{C}_\pi, \pi \in \mathcal{P}$ is a collection of congruent cones (i.e., cones with the same angular measure) that symmetrically partition $\mathbb{R}^n$. Moreover, for every pair $(\tau, \pi) \in \mathcal{P} \times \mathcal{P}$ there exists a permutation matrix $P_{\tau,\pi}$ such that $P_{\tau,\pi}\mathcal{C}_\tau = \mathcal{C}_\pi$ and

$$\|\mathbf{x} - P_{\pi,\tau}\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|, \, \mathbf{x} \in \mathcal{C}_\tau, \mathbf{y} \in \mathcal{C}_\pi. \tag{A.38}$$

The above inequality follows because of the three following facts: (i) $P_{\tau,\pi}\mathcal{C}_\tau = \mathcal{C}_\pi$ implies that $\mathcal{C}_\pi$ is a reflection of $\mathcal{C}_\tau$ along some hyperplane $\mathcal{T}$; (ii) the hyperplane $\mathcal{T}$ bisects the distance between $P_{\pi,\tau}\mathbf{y}$ and $\mathbf{y}$ into equal segments; and (iii) $\mathbf{x}$ and $P_{\pi,\tau}\mathbf{y}$ are on the same side of the hyperplane and $\mathbf{y}$ is on the opposite side of the hyperplane. Thus, the distance between $\mathbf{x}$ and $P_{\pi,\tau}\mathbf{y}$ is smaller than the distance between $\mathbf{x}$ and $\mathbf{y}$.

Next, with some abuse of notation, we let $f_{\mathbf{Z}}(\|\mathbf{z}\|)$ denote the PDF of $\mathbf{Z}$. This notation highlights the fact that the PDF of $\mathbf{Z}$ only depends on the norm. We also define $\boldsymbol{\mu} = \tilde{K}^{-\frac{1}{2}}\tilde{\mathbf{y}}$

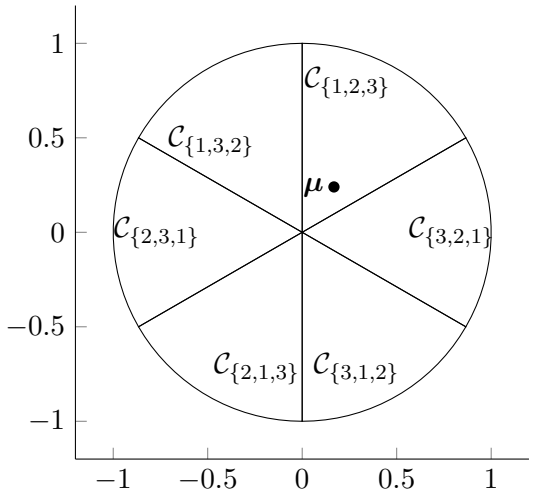


Figure A.1: A pictorial depiction of the inequality in (A.39) for $n = 3$ and $\tau = \{1, 2, 3\}$.

where $\boldsymbol{\mu} \in \mathcal{C}_\tau$ since by assumption $\tilde{\mathbf{y}} \in \mathcal{H}_\tau$. With this, we obtain

$$\begin{aligned}
\Pr\left(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\pi\right) &= \Pr\left(\mathbf{Z} + \left(K_{\mathbf{N}}^{-1} + I_n\right)^{\frac{1}{2}} \tilde{\mathbf{y}} \in \left(K_{\mathbf{N}}^{-1} + I_n\right)^{\frac{1}{2}} \mathcal{H}_\pi\right) \\
&\overset{(a)}{=} \Pr(\mathbf{Z} + \boldsymbol{\mu} \in \mathcal{C}_\pi) = \int_{\mathcal{C}_\pi} f_{\mathbf{Z}}(\|\mathbf{z} - \boldsymbol{\mu}\|)\, \mathrm{d}\mathbf{z} \\
&\overset{(b)}{\leq} \int_{\mathcal{C}_\pi} f_{\mathbf{Z}}(\|P_{\pi,\tau}\mathbf{z} - \boldsymbol{\mu})\|)\, \mathrm{d}\mathbf{z} \\
&\overset{(c)}{=} \int_{P_{\pi,\tau}\mathcal{C}_\pi} f_{\mathbf{Z}}(\|\mathbf{z} - \boldsymbol{\mu})\|)\, \mathrm{d}\mathbf{z} \\
&\overset{(d)}{=} \int_{\mathcal{C}_\tau} f_{\mathbf{Z}}(\|\mathbf{z} - \boldsymbol{\mu})\|)\, \mathrm{d}\mathbf{z} \\
&= \Pr\left(\mathbf{Z} + \boldsymbol{\mu} \in \mathcal{C}_\tau\right) \\
&= \Pr\left(\tilde{\mathbf{Y}}_0 + \tilde{\mathbf{y}} \in \mathcal{H}_\tau\right), \qquad \text{(A.39)}
\end{aligned}$$

where the labeled (in)equalities follow from: (a) letting $\boldsymbol{\mu} = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{1/2} \tilde{\mathbf{y}}$ and remembering that $\mathcal{C}_\pi = \tilde{K}^{-\frac{1}{2}} \mathcal{H}_\pi = \left(K_{\mathbf{N}}^{-1} + I_n\right)^{1/2} \mathcal{H}_\pi$ for all $\pi \in \mathcal{P}$; (b) applying the bound in (A.38) and noting that $\boldsymbol{\mu} \in \mathcal{C}_\tau$; (c) using change of variable and the fact that $|\det(P_{\tau,\pi})| = 1$; and (d) the fact that $\mathcal{C}_\tau = P_{\pi,\tau}\mathcal{C}_\pi$. The geometric interpretation of the inequality in (b) is shown in Fig. A.1. In particular, in Fig. A.1 the view is taken with respect to the axis of symmetry. The

dashed ball centered at $\boldsymbol{\mu}$ is meant to represent a level set of the PDF of $\mathbf{Z}+\boldsymbol{\mu}$. The intersection of the dashed ball and a cone $\mathcal{C}_\pi$ is the largest for the cone in which $\boldsymbol{\mu}$ lies, i.e., $\pi = \{1,2,3\}$. The proof of Lemma 3.5.9 is concluded by noting that (A.39) holds with equality if $\tau = \pi$.

A.9 Proof of Lemma A.6.1

Let $\mathcal{E}^n$ be an n-dimensional ellipsoid centered at the origin with unitary axes $\{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_n\}$ and corresponding radii equal to $\{r_1, r_2, \ldots, r_n\}$. Let one of the axes of $\mathcal{E}^n$ be equal to $\frac{1}{\sqrt{n}}\mathbf{1}_n$. Specifically, without loss of generality, we set $\boldsymbol{\nu}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n$, which has r_n as the corresponding radius. Then, by introducing the hyperplane $\mathcal{W}(r) = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\nu}_n^T \mathbf{x} = r\}$, for any $\pi \in \mathcal{P}$, we can represent the volume of the intersection between $\mathcal{H}_\pi$ and $\mathcal{E}^n$ as

$$\begin{aligned}\mathrm{Vol}^n\left(\mathcal{H}_\pi \cap \mathcal{E}^n\right) &= \int_{-r_n}^{r_n} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}^n \cap \mathcal{W}(r)\right) \mathrm{d}r \\ &= \int_{-r_n}^{r_n} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}^{n-1}_{\mathcal{W}(r)}\right) \mathrm{d}r, \end{aligned} \tag{A.40}$$

where $\mathcal{E}^{n-1}_{\mathcal{W}(r)} = \mathcal{E}^n \cap \mathcal{W}(r)$ is an $(n-1)$-dimensional ellipsoid in $\mathbb{R}^n$.

Note that since $\mathcal{E}^{n-1}_{\mathcal{W}(r)}$ has $\boldsymbol{\nu}_n$ as normal vector, which is one of the axes of $\mathcal{E}^n$, the ellipsoid $\mathcal{E}^{n-1}_{\mathcal{W}(r)}$ can be represented as

$$\mathcal{E}^{n-1}_{\mathcal{W}(r)} = m(r) I_n \cdot \mathcal{E}^{n-1}_{\mathcal{W}(0)} + r\boldsymbol{\nu}_n, \tag{A.41}$$

where $m(r) : [-r_n, r_n] \to (0, 1]$ is some magnitude function. Then, we have

$$\begin{aligned}\mathrm{Vol}^n\left(\mathcal{H}_\pi \cap \mathcal{E}^n\right) &\overset{(a)}{=} \int_{-r_n}^{r_n} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \left\{m(r) I_n \cdot \mathcal{E}^{n-1}_{\mathcal{W}(0)} + r\boldsymbol{\nu}_n\right\}\right) \mathrm{d}r \\ &\overset{(b)}{=} \int_{-r_n}^{r_n} \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap m(r) I_n \cdot \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) \mathrm{d}r \\ &\overset{(c)}{=} \int_{-r_n}^{r_n} \left|\det\left(m(r) I_n\right)\right| \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) \mathrm{d}r \\ &= \mathrm{Vol}^{n-1}\left(\mathcal{H}_\pi \cap \mathcal{E}^{n-1}_{\mathcal{W}(0)}\right) \int_{-r_n}^{r_n} m(r)^n \mathrm{d}r, \end{aligned} \tag{A.42}$$

where the labeled equalities follow from: (a) substituting (A.41) into (A.40); (b) the fact that $\mathcal{H}_\pi,\ \forall \pi \in \mathcal{P}$ is invariant to adding $a\boldsymbol{\nu}_n$, where $a \in \mathbb{R}$ is any constant and remember that

$\boldsymbol{\nu}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n$ (i.e., $\mathcal{H}_\pi = \mathcal{H}_\pi + a\boldsymbol{\nu}_n$); and (c) the facts that, for any invertible matrix A and any set $\mathcal{S}$, $\mathrm{Vol}^n(A\mathcal{S}) = |\det(A)|\mathrm{Vol}^n(\mathcal{S})$ and $\mathcal{H}_\pi = kI_n\mathcal{H}_\pi$, where k is any positive number. We conclude the proof of Lemma A.6.1 by defining $c(r_n) = \int_{-r_n}^{r_n} m(r)^n \mathrm{d}r$.

A.10 Eigenvalues of B in (A.32)

We seek to compute the eigenvalues of the matrix B defined as

$$B = \begin{bmatrix} \gamma I_{n-2} & 0_{(n-2)\times 2} \\ 0_{2\times(n-2)} & S \end{bmatrix}, \tag{A.43}$$

where $S = \left[\begin{smallmatrix} \gamma & v \\ v & a \end{smallmatrix}\right]$ is a 2×2 symmetric matrix. These can be found as the values of λ that satisfy the equation

$$\begin{aligned} \det(B - \lambda I_n) = 0 &\implies \det\left(\begin{bmatrix} (\gamma-\lambda)I_{n-2} & 0_{(n-2)\times 2} \\ 0_{2\times(n-2)} & S - \lambda I_2 \end{bmatrix}\right) = 0 \\ &\implies \det\left((\gamma-\lambda)I_{n-2}\right)\det\left(S - \lambda I_2\right) = 0 \\ &\implies (\gamma-\lambda)^{n-2}((a-\lambda)(\gamma-\lambda) - v^2) = 0. \end{aligned}$$

Hence the matrix B in (A.43) has $n-2$ eigenvalues equal to γ and the remaining two eigenvalues can be found as the solution of

$$(a-\lambda)(\gamma-\lambda) - v^2 = 0 \implies \lambda = \frac{a+\gamma \pm \sqrt{(a-\gamma)^2+4v^2}}{2}. \tag{A.44}$$

Appendix B

Differed Proofs in Chapter 4

B.1 Proof of Lemma 4.3.5

We start by noting that, for an exchangeable $\mathbf{X} \in \mathbb{R}^n$, the optimal decision regions (2.4) are

$$\mathcal{R}_{\tau,K_{\mathbf{N}}} = \left\{ \mathbf{y} \in \mathbb{R}^n : f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau) > \max_{\substack{\eta \in \mathcal{P} \\ \eta \neq \tau}} f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\eta) \right\}. \tag{B.1}$$

Since $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, we have that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}|\mathcal{H}_\tau) &= \int_{\mathbf{x} \in \mathbb{R}^n} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x} \mid \mathcal{H}_\tau) \, \mathrm{d}\mathbf{x} \\ &= \frac{1}{n!} \int_{\mathbf{x} \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \end{aligned} \tag{B.2}$$

where the last equality follows since $\mathbf{X}$ is exchangeable, and thus $f_{\mathbf{X}}(\mathbf{x} \mid \mathcal{H}_\tau) = \frac{1}{n!} \mathbb{1}\{\mathbf{x} \in \mathcal{H}_\tau\} f_{\mathbf{X}}(\mathbf{x})$. Moreover, due to the exchangeability of $\mathbf{X}$, for any permutation matrix P, we have

that $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(P\mathbf{x})$. Let $\mathbf{y} \in \mathcal{H}_\tau$; then, for any $\eta \in \mathcal{P}$, we have that

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y} \mid \mathcal{H}_\tau) - f_{\mathbf{Y}}(\mathbf{y} \mid \mathcal{H}_\eta) &= \frac{1}{n!} \int_{\mathbf{x} \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\quad - \frac{1}{n!} \int_{\mathbf{u} \in \mathcal{H}_\eta} f_{\mathbf{N}}(\mathbf{y} - \mathbf{u}) f_{\mathbf{X}}(\mathbf{u}) \, \mathrm{d}\mathbf{u} \\ &\overset{(\mathrm{a})}{=} \frac{1}{n!} \int_{\mathbf{x} \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &\quad - \frac{1}{n!} \int_{\mathbf{x} \in \mathcal{H}_\tau} f_{\mathbf{N}}(\mathbf{y} - P_{\tau,\eta}\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\ &= \frac{1}{n!} \int_{\mathbf{x} \in \mathcal{H}_\tau} (f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) - f_{\mathbf{N}}(\mathbf{y} - P_{\tau,\eta}\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \end{aligned} \tag{B.3}$$

where the equality in (a) is due to the change of variable where $P_{\tau,\eta}$ is the permutation matrix that permutes τ in η. To finalize the proof we utilize [1, Lemma 1], which we restate below for completeness.

Lemma B.1.1 (Lemma 1 in [1])**.** *Let $\mathbf{u} \in \mathcal{H}_\tau$ and $\mathbf{w} \in \mathcal{H}_\tau$. Then, for any permutation matrix P, we have that*

$$\|\mathbf{u} - \mathbf{w}\|_2 \leq \|\mathbf{u} - P\mathbf{w}\|_2. \tag{B.4}$$

Since $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, then $f_{\mathbf{N}}(\mathbf{y} - \mathbf{x})$ and $f_{\mathbf{N}}(\mathbf{y} - P_{\tau,\eta}\mathbf{x})$ in (B.3) are monotonically decreasing with respect to $\|\mathbf{y} - \mathbf{x}\|_2$ and $\|\mathbf{y} - P_{\tau,\eta}\mathbf{x}\|_2$, respectively. By Lemma B.1.1, we then have that, for any $\eta \in \mathcal{P}$,

$$f_{\mathbf{N}}(\mathbf{y} - \mathbf{x}) - f_{\mathbf{N}}(\mathbf{y} - P_{\tau,\eta}\mathbf{x}) \geq 0, \tag{B.5}$$

which implies that (B.3) is non-negative. Hence, we have that if $\mathbf{y} \in \mathcal{H}_\tau$,

$$f_{\mathbf{Y}}(\mathbf{y} \mid \mathcal{H}_\tau) \geq f_{\mathbf{Y}}(\mathbf{y} \mid \mathcal{H}_\eta), \ \forall \eta \in \mathcal{P}. \tag{B.6}$$

The optimal decision regions in (B.1) are then given by

$$\mathcal{R}_{\tau, K_{\mathbf{N}}} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \in \mathcal{H}_\tau\} = \mathcal{H}_\tau. \tag{B.7}$$

By noting that $\mathcal{H}_\tau = c\mathcal{H}_\tau, \ \forall c > 0$, we conclude the proof of Lemma 4.3.5.

B.2 Proof of Theorem 4.4.1

Before proceeding with the proof of Theorem 4.4.1, we present an ancillary result.

Lemma B.2.1. *Let* $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_{n-1}, \tilde{K})$ *where* $\tilde{K}$ *is defined in* (4.22)*. Then, for any subset* $\mathcal{I} \subseteq [1 : n-1]$*,*

$$\Pr\left(\bigcap_{i\in\mathcal{I}}\{V_i \le t_i\}\right) \le \prod_{i\in\mathcal{I}} \Pr\left(\{V_i \le t_i\}\right). \tag{B.8}$$

Proof. The bound in (B.8) holds if the random vector $\mathbf{V}$ consists of *negatively associated* random variables [85]. Observe that the Gaussian random vector $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_{n-1}, \tilde{K})$ consists of either negatively correlated or independent random variables (see the structure of $\tilde{K}$ in (4.22)). As it was shown in [85], this implies that the random variables in $\mathbf{V}$ are negatively associated. This concludes the proof of Lemma B.2.1. □

From Theorem 4.3.2 with $K_{\mathbf{N}} = \sigma^2 I_n$, $A = I_n$ and $\mathbf{b} = \mathbf{0}_n$, we have that

$$\begin{aligned} P_{e,\text{lin}}(\sigma) &= 1 - \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[Q_{\sigma^2\tilde{K}}\left(-T_\tau \mathbf{X}\right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right] P_{\mathbf{X}}(\mathcal{H}_\tau) \\ &= 1 - \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[Q_{\sigma^2\tilde{K}}\left(-\mathbf{W}_\tau\right)\right] P_{\mathbf{X}}(\mathcal{H}_\tau), \end{aligned} \tag{B.9}$$

where in the first equality $\tilde{K}$ is given by (4.22), $Q_{\sigma^2\tilde{K}}(\cdot)$ is the multivariate Gaussian Q-function with covariance matrix $\sigma^2\tilde{K}$, and in the second equality $\mathbf{W}_\tau$ is the τ-spacing of $\mathbf{X}$ in Definition 4.2.2.

Letting $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_{n-1}, \tilde{K})$, we then have

$$\begin{aligned} P_{e,\text{lin}}(\sigma) =& 1 - \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[\Pr\left(\mathbf{V} \ge -\frac{\mathbf{W}_\tau}{\sigma}\right)\right] P_{\mathbf{X}}(\mathcal{H}_\tau) \\ =& \sum_{\tau\in\mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau) - \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[\Pr\left(\mathbf{V} \ge -\frac{\mathbf{W}_\tau}{\sigma}\right)\right] P_{\mathbf{X}}(\mathcal{H}_\tau) \\ =& \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[1 - \Pr\left(\mathbf{V} \ge -\frac{\mathbf{W}_\tau}{\sigma}\right)\right] P_{\mathbf{X}}(\mathcal{H}_\tau). \end{aligned} \tag{B.10}$$

The expectation in (B.10) can be equivalently written as

$$\begin{aligned}\mathbb{E}\left[1-\Pr\left(\bigcap_{i=1}^{n-1}\left\{V_i \geq -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right\}\right)\right] &= \mathbb{E}\left[\Pr\left(\bigcup_{i=1}^{n-1}\left\{V_i < -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right\}\right)\right] \\ &= \Pr\left(\bigcup_{i=1}^{n-1}\left\{V_i < -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right\}\right) \\ &= \sum_{k=1}^{n-1}\left((-1)^{k-1}\sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=k}}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right)\right), \qquad \text{(B.11)}\end{aligned}$$

where the last equality follows from the inclusion-exclusion principle where $\mathcal{A}_{\mathcal{I}}^{\tau} = \cap_{i\in\mathcal{I}}\mathcal{A}_i^{\tau}$ with $\mathcal{A}_i^{\tau} = \left\{V_i < -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right\}$. From the expression in (B.11) it follows that

$$\lim_{\sigma\to 0}\frac{P_{e,\text{lin}}(\sigma)}{\sigma} = \sum_{\tau\in\mathcal{P}}\sum_{k=1}^{n-1}\left((-1)^{k-1}\sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=k}}\lim_{\sigma\to 0}\frac{1}{\sigma}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right)\right)P_{\mathbf{X}}(\mathcal{H}_\tau). \qquad \text{(B.12)}$$

In what follows, we therefore analyze

$$\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right) = \Pr\left(\bigcap_{i\in\mathcal{I}}\left\{V_i < -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right\}\right), \qquad \text{(B.13)}$$

by considering two separate cases.

• **Case 1:** $k = 1$. Let $\mathcal{I} = \{i\}$; then, we can write (B.13) as

$$\begin{aligned}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right) &= \mathbb{E}\left[\Pr\left(V_i < -\frac{(\mathbf{W}_\tau)_i}{\sigma}\right)\right] \\ &\overset{\text{(a)}}{=} \mathbb{E}\left[Q\left(\frac{(\mathbf{W}_\tau)_i}{\sqrt{2}\sigma}\right)\right] \\ &= \int_0^\infty Q\left(\frac{w}{\sqrt{2}\sigma}\right) f_{(\mathbf{W}_\tau)_i}(w)\,\mathrm{d}w \\ &\overset{\text{(b)}}{=} \int_0^\infty Q\left(u\right) f_{(\mathbf{W}_\tau)_i}(\sqrt{2}\sigma u)\sqrt{2}\sigma\,\mathrm{d}u, \qquad \text{(B.14)}\end{aligned}$$

where the labeled equalities follow from: (a) the fact that $V_i \sim \mathcal{N}(0,2)$; and (b) applying a

change of variable. Thus, we have that

$$\begin{aligned}\lim_{\sigma\to 0}\frac{1}{\sigma}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right) &= \sqrt{2}\int_0^{\infty}\lim_{\sigma\to 0} Q\left(u\right) f_{(\mathbf{W}_\tau)_i}(\sqrt{2}\sigma u)\,\mathrm{d}u \\ &= \sqrt{2} f_{(\mathbf{W}_\tau)_i}\left(0^+\right)\int_0^{\infty} Q\left(u\right)\mathrm{d}u \\ &= \frac{f_{(\mathbf{W}_\tau)_i}\left(0^+\right)}{\sqrt{\pi}}, \end{aligned} \tag{B.15}$$

where the first equality follows from the dominated convergence theorem, which is verifiable since for any σ, $Q(u)f_{(\mathbf{W}_\tau)_i}(\sqrt{2}\sigma u) \leq Q(u)\max_w f_{(\mathbf{W}_\tau)_i}(w) < \infty$ due to the assumption $\max_w f_{(\mathbf{W}_\tau)_i}(w) < \infty,\ \forall i$. □

• **Case 2:** $k \geq 2$. By using the bound in Lemma B.2.1, we obtain

$$\begin{aligned}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right) &\leq \mathbb{E}\left[\prod_{i\in\mathcal{I}}\Pr\left(V_i < \frac{-(\mathbf{W}_\tau)_i}{\sigma}\right)\right] \\ &= \mathbb{E}\left[\prod_{i\in\mathcal{I}} Q\left(\frac{(\mathbf{W}_\tau)_i}{\sqrt{2}\sigma}\right)\right] \\ &\leq \mathbb{E}\left[\prod_{\substack{i\in\mathcal{J}\\ \mathcal{J}\subset\mathcal{I},|\mathcal{J}|=2}} Q\left(\frac{(\mathbf{W}_\tau)_i}{\sqrt{2}\sigma}\right)\right]. \end{aligned} \tag{B.16}$$

By letting $\mathcal{J} = \{s,t\} \subset \mathcal{I}$ in (B.16), we obtain that

$$\begin{aligned}\lim_{\sigma\to 0}\frac{\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right)}{\sigma} &\leq \lim_{\sigma\to 0}\sigma\int_0^{\infty}\int_0^{\infty} Q\left(\frac{w}{\sqrt{2}}\right) Q\left(\frac{z}{\sqrt{2}}\right) f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(\sigma w,\sigma z)\mathrm{d}w\mathrm{d}z \\ &\leq \lim_{\sigma\to 0}\sigma\frac{f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(0^+,0^+)}{\pi} \\ &= 0, \end{aligned} \tag{B.17}$$

where the equality follows from the assumption that $\max_{u,v} f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(u,v) < \infty,\ \forall s,t,$

and the second inequality is due to the fact that

$$\begin{aligned}
&\lim_{\sigma\to 0}\int_0^\infty\int_0^\infty Q\left(\frac{w}{\sqrt{2}}\right)Q\left(\frac{z}{\sqrt{2}}\right)f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(\sigma w,\sigma z)\mathrm{d}w\mathrm{d}z\\
&\overset{(a)}{=}\int_0^\infty\int_0^\infty Q\left(\frac{w}{\sqrt{2}}\right)Q\left(\frac{z}{\sqrt{2}}\right)f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(0^+,0^+)\mathrm{d}w\mathrm{d}z\\
&=\frac{f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(0^+,0^+)}{\pi}<\infty,
\end{aligned}\tag{B.18}$$

where (a) follows from the dominated convergence theorem, which is verifiable by means of the assumption $\max_{u,v} f_{(\mathbf{W}_\tau)_s,(\mathbf{W}_\tau)_t}(u,v)<\infty,\ \forall s,t$ (similar to Case 1). □

By using the limits in (B.15) and (B.17) inside (B.12), we obtain

$$\begin{aligned}
\lim_{\sigma\to 0}\frac{P_{e,\mathrm{lin}}(\sigma)}{\sigma}&\overset{(a)}{=}\sum_{\tau\in\mathcal{P}}\sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=1}}\lim_{\sigma\to 0}\frac{1}{\sigma}\Pr\left(\mathcal{A}_{\mathcal{I}}^{\tau}\right)P_{\mathbf{X}}(\mathcal{H}_\tau)\\
&\overset{(b)}{=}\sum_{\tau\in\mathcal{P}}\sum_{i=1}^{n-1}P_{\mathbf{X}}(\mathcal{H}_\tau)\frac{f_{(\mathbf{W}_\tau)_i}(0^+)}{\sqrt{\pi}},
\end{aligned}\tag{B.19}$$

where (a) follows from (B.17), and (b) follows from (B.15). Equivalently, (B.17) and (B.19) imply that in the low-noise regime

$$P_{e,\mathrm{lin}}(\sigma)=\sum_{\tau\in\mathcal{P}}\sum_{i=1}^{n-1}P_{\mathbf{X}}(\mathcal{H}_\tau)\frac{f_{(\mathbf{W}_\tau)_i}(0^+)}{\sqrt{\pi}}\sigma+O(\sigma^2).\tag{B.20}$$

This concludes the proof of Theorem 4.4.1.

B.3 Proof of Proposition 4.4.4

We first state the following lemma, the proof of which can be found in Appendix B.8.2.

Lemma B.3.1. *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*

$$\|f_X\|_2<\infty.\tag{B.21}$$

Then, the following holds

$$f_{W_i}(u) < \infty,\ \forall u \in \mathbb{R}_+, 1 \le i \le n-1, \text{ and} \tag{B.22}$$

$$f_{W_i,W_j}(u,v) < \infty,\ \forall (u,v) \in \mathbb{R}_+^2,\ 1 \le i < j \le n-1, \tag{B.23}$$

where W_i is the i-th spacing of $\mathbf{X}$ in Definition 4.2.1.

Lemma B.3.1 ensures that $\mathbf{X}$ with the assumption in (4.31) satisfies the assumption in (4.28) in Corollary 4.4.3. Then, from Corollary 4.4.3, we have that

$$\lim_{\sigma\to 0} \frac{P_{e,\mathrm{lin}}(\sigma)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}. \tag{B.24}$$

Now, to complete the proof we show that

$$\sum_{i=1}^{n-1} f_{W_i}(0^+) = n(n-1)\int_{-\infty}^{\infty} f_X^2(x)\,\mathrm{d}x. \tag{B.25}$$

For i.i.d. $X_i \sim F_X$, where F_X is the cdf of X, we observe that from [60] and with $\beta_{n,i} := \frac{n!}{(i-1)!(n-i-1)!}$ we have

$$\begin{aligned}
\sum_{i=1}^{n-1} f_{W_i}(0^+) &= \sum_{i=1}^{n-1} \beta_{n,i} \int_{-\infty}^{\infty} (F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&\overset{(a)}{=} \int_{-\infty}^{\infty} \sum_{i=1}^{n-1} \beta_{n,i} (F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sum_{i=1}^{n-1} \binom{n}{i-1} (n-i+1)(n-i)(F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&\overset{(b)}{=} \int_{-\infty}^{\infty} \sum_{j=0}^{n-2} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^{j}(1-F_X(x))^{n-j-2} f_X^2(x)\,\mathrm{d}x,
\end{aligned} \tag{B.26}$$

where (a) follows by using the Fubini-Tonelli theorem, and (b) follows from the change of

variable $j = i - 1$. To simplify the integrand in (B.26), we make use of the following,

$$\begin{aligned}
&\sum_{j=0}^{n-2} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^j (1-F_X(x))^{n-j-2} \\
&= \sum_{j=0}^{n} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^j (1-F_X(x))^{n-j-2} \\
&\overset{(a)}{=} \mathbb{E}\left[(n-B)(n-B-1)\right](1-F_X(x))^{-2} \\
&= \mathbb{E}\left[n^2 - 2nB + B^2 - n + B\right](1-F_X(x))^{-2} \\
&= \Big(n^2 - 2n^2 F_X(x) + nF_X(x)(1-F_X(x)) \\
&\qquad + n^2 F_X^2(x) - n + nF_X(x)\Big)(1-F_X(x))^{-2} \\
&= n(n-1)(1-2F_X(x)+F_X^2(x))(1-F_X(x))^{-2} \\
&= n(n-1), \qquad \text{(B.27)}
\end{aligned}$$

where in (a) we let $B \sim \mathrm{Bin}(n, F_X(x))$ be the binomial random variable with parameters n and $F_X(x)$, and the expectation is with respect to B.

With this, we obtain that (B.26) reduces to

$$\begin{aligned}
\sum_{i=1}^{n-1} f_{W_i}(0^+) &= \int_{-\infty}^{\infty} n(n-1) f_X^2(x)\, \mathrm{d}x \\
&= n(n-1) \int_{-\infty}^{\infty} f_X^2(x)\, \mathrm{d}x, \qquad \text{(B.28)}
\end{aligned}$$

and thus,

$$\lim_{\sigma \to 0} \frac{P_{e,\mathrm{lin}}(\sigma)}{\sigma} = \frac{n(n-1)}{\sqrt{\pi}} \int_{-\infty}^{\infty} f_X^2(x)\, \mathrm{d}x. \qquad \text{(B.29)}$$

This concludes the proof of Proposition 4.4.4.

B.4 Proof of Theorem 4.5.1

We start by noting that, in view of the limit in (4.35), we have that $\lim_{\sigma\to\infty} P_{c,\text{lin}} = \frac{1}{n!}$. We now consider the following limit,

$$\lim_{\sigma\to\infty} \frac{P_{c,\text{lin}} - \frac{1}{n!}}{\frac{1}{\sigma}}. \tag{B.30}$$

Instead of working with σ, we parameterize the problem in terms of $\sigma = \frac{1}{\kappa}$. Then, (B.30) can be equivalently expressed as

$$\lim_{\sigma\to\infty} \frac{P_{c,\text{lin}} - \frac{1}{n!}}{\frac{1}{\sigma}} = \lim_{\kappa\to 0} \frac{P_{c,\text{lin}} - \frac{1}{n!}}{\kappa} = \lim_{\kappa\to 0} \frac{\partial P_{c,\text{lin}}}{\partial \kappa}, \tag{B.31}$$

where the last equality can be argued by using the definition of the derivative or L'Hôpital's rule. We make use of the following lemma, proved in Appendix B.8.3, to prove the Theorem 4.5.1.

Lemma B.4.1. *Suppose $\kappa < \infty$. Then, for any constants a and $c_i,\ i \in [1:n]$, the following identity holds*

$$\begin{aligned}
&\frac{\partial}{\partial \kappa} \int_{c_1\kappa}^{a} \cdots \int_{c_n\kappa}^{a} f(\mathbf{x})\, \mathrm{d}x_n \cdots\, \mathrm{d}x_1 \\
&= -\sum_{i=1}^{n} c_i \int_{c_1\kappa}^{a} \cdots \int_{c_n\kappa}^{a} f(\mathbf{x})|_{x_i = c_i\kappa}\, \mathrm{d}x_n \cdots \mathrm{d}x_{i+1} \mathrm{d}x_{i-1} \cdots \mathrm{d}x_1.
\end{aligned} \tag{B.32}$$

We now note that from Theorem 4.3.2, the probability of correctness under $\phi_{\text{lin}}(\cdot; I_n, \mathbf{0}_n)$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$ is given by

$$\begin{aligned}
P_{c,\text{lin}} &= 1 - P_{e,\text{lin}} \\
&= \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[Q_{\sigma^2\tilde{K}}\left(-T_\tau \mathbf{X}\right) \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right] P_{\mathbf{X}}(\mathcal{H}_\tau) \\
&= \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[Q_{\sigma^2\tilde{K}}\left(-\mathbf{W}_\tau\right)\right] P_{\mathbf{X}}(\mathcal{H}_\tau),
\end{aligned} \tag{B.33}$$

where in the first equality $\tilde{K}$ is given by (4.22), $Q_{\sigma^2\tilde{K}}(\cdot)$is the multivariate Gaussian Q-function with covariance matrix $\sigma^2\tilde{K}$, and in the second equality $\mathbf{W}_\tau$ is the τ-spacing of $\mathbf{X}$ in Definition 4.2.2.

By Lemma B.4.1 and letting $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_n, \tilde{K})$, the derivative of $P_{c,\text{lin}}$ with respect to κ is

given by

$$\frac{\partial P_{c,\text{lin}}}{\partial \kappa} \overset{(a)}{=} \sum_{\tau\in\mathcal{P}} \mathbb{E}\left[\frac{\partial}{\partial\kappa}\int_{\mathbf{v}\geq -\kappa\mathbf{W}_\tau} f_{\mathbf{V}}(\mathbf{v})\,\mathrm{d}\mathbf{v}\right] P_{\mathbf{X}}(\mathcal{H}_\tau)$$
$$\overset{(b)}{=} \sum_{\tau\in\mathcal{P}} \mathbb{E}\Bigg[\sum_{i=1}^{n-1}(\mathbf{W}_\tau)_i$$
$$\int_{-(\mathbf{W}_\tau)_1\kappa}^{\infty}\cdots\int_{-(\mathbf{W}_\tau)_n\kappa}^{\infty} f_{\mathbf{V}_{\setminus i},V_i}(\mathbf{v}_{\setminus i},-(\mathbf{W}_\tau)_i\kappa)\mathrm{d}\mathbf{v}_{\setminus i}\Bigg] P_{\mathbf{X}}(\mathcal{H}_\tau), \tag{B.34}$$

where in (a) we used the Leibniz's integral rule, and (b) follows by Lemma B.4.1, where we let $f_{\mathbf{V}_{\setminus i},V_i}(\mathbf{v}_{\setminus i},-(\mathbf{W}_\tau)_i\kappa) = f_{\mathbf{V}}(\mathbf{v})|_{v_i=-(\mathbf{W}_\tau)_i\kappa}$ and $\mathbf{V}_{\setminus i}$ is an $(n-1)$-dimensional vector obtained by retaining all the entries of $\mathbf{V}$ except for the i-th one, i.e., $\mathbf{V}_{\setminus i} = [V_1,...,V_{i-1},V_{i+1},...,V_n]^T$.

By taking the limit of (B.34), we get (B.31). In particular, we obtain

$$\lim_{\kappa\to 0}\frac{\partial P_{c,\text{lin}}}{\partial\kappa} \overset{(a)}{=} \sum_{\tau\in\mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau)\mathbb{E}\left[\sum_{i=1}^{n-1}(\mathbf{W}_\tau)_i\int_0^\infty\cdots\int_0^\infty f_{\mathbf{V}_{\setminus i},V_i}(\mathbf{v}_{\setminus i},0)\mathrm{d}\mathbf{v}_{\setminus i}\right]$$
$$= \sum_{\tau\in\mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau)\sum_{i=1}^{n-1}\mathbb{E}\left[(\mathbf{W}_\tau)_i\right]\int_{\mathbb{R}_+^{n-2}} f_{\mathbf{V}_{\setminus i},V_i}(\mathbf{v}_{\setminus i},0)\mathrm{d}\mathbf{v}_{\setminus i}$$
$$\overset{(b)}{=} \sum_{\tau\in\mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau)\sum_{i=1}^{n-1}\mathbb{E}[(\mathbf{W}_\tau)_i]\Pr(\mathbf{V}_{\setminus i}\geq\mathbf{0}_{n-2}\,|\,V_i=0)f_{V_i}(0), \tag{B.35}$$

where the labeled equalities follow from: (a) using the dominated convergence theorem, which is verified since

$$\sum_{i=1}^{n-1}(\mathbf{W}_\tau)_i\int_{-(\mathbf{W}_\tau)_1\kappa}^{\infty}\cdots\int_{-(\mathbf{W}_\tau)_n\kappa}^{\infty} f_{\mathbf{V}_{\setminus i},V_i}(\mathbf{v}_{\setminus i},-(\mathbf{W}_\tau)_i\kappa)\mathrm{d}\mathbf{v}_{\setminus i}$$
$$\leq \sum_{i=1}^{n-1}(\mathbf{W}_\tau)_i f_{V_i}(-(\mathbf{W}_\tau)_i\kappa), \tag{B.36}$$

and $\mathbb{E}[(\mathbf{W}_\tau)_i]$ is bounded due to the assumption $\mathbb{E}[|X_i|] < \infty$ for all $i \in [1:n]$, i.e.,

$$\begin{aligned}\mathbb{E}[(\mathbf{W}_\tau)_i] &= \mathbb{E}[X_{\tau_{i+1}} - X_{\tau_i} \mid \mathbf{X} \in \mathcal{H}_\tau] \\ &= \mathbb{E}[X_{\tau_{i+1}} \mid \mathbf{X} \in \mathcal{H}_\tau] - \mathbb{E}[X_{\tau_i} \mid \mathbf{X} \in \mathcal{H}_\tau] \\ &\leq \mathbb{E}[|X_{\tau_{i+1}}| \mid \mathbf{X} \in \mathcal{H}_\tau] + \mathbb{E}[|X_{\tau_i}| \mid \mathbf{X} \in \mathcal{H}_\tau] \\ &< \infty, \end{aligned} \tag{B.37}$$

and $f_{V_i}(\cdot)$ is a Gaussian density function that is bounded; and (b) using the following,

$$\begin{aligned}\int_{\mathbb{R}_+^{n-2}} f_{\mathbf{V}_{\setminus i}, V_i}(\mathbf{v}_{\setminus i}, 0)\, \mathrm{d}\mathbf{v}_{\setminus i} &= \int_{\mathbb{R}_+^{n-2}} f_{\mathbf{V}_{\setminus i}|V_i}(\mathbf{v}_{\setminus i} \mid 0) f_{V_i}(0) \mathrm{d}\mathbf{v}_{\setminus i} \\ &= \Pr(\mathbf{V}_{\setminus i} \geq \mathbf{0}_{n-2} | V_i = 0) f_{V_i}(0). \end{aligned} \tag{B.38}$$

To finalize the proof, it remains to compute $\Pr(\mathbf{V}_{\setminus i} \geq \mathbf{0}_{n-2} | V_i = 0) = \alpha_i$. This is done as

follows,

$$
\begin{aligned}
\alpha_i &= \Pr(\mathbf{V}_{\setminus i} \geq \mathbf{0}_{n-2} \mid V_i = 0) \\
&\overset{(a)}{=} \Pr\left(\bigcap_{j=1, j\neq i}^{n-1} \{Z_{j+1} - Z_j \geq 0\} \,\middle|\, Z_{i+1} - Z_i = 0 \right) \\
&= \Pr\left(\{Z_1 \leq \cdots \leq Z_i\} \cap \{Z_{i+1} \leq \cdots \leq Z_n\} \mid Z_{i+1} = Z_i \right) \\
&\overset{(b)}{=} \mathbb{E}\left[\Pr\left(\{\cdots \leq Z_i\} \cap \{Z_{i+1} \leq \cdots\} | Z_{i+1}, Z_i \right) | Z_{i+1} = Z_i \right] \\
&= \int_{-\infty}^{\infty} \Pr\left(\{\cdots \leq t\} \cap \{t \leq \cdots\} \right) f_{Z_i, Z_{i+1} | Z_i = Z_{i+1}}(t,t) \mathrm{d}t \\
&\overset{(c)}{=} \int_{-\infty}^{\infty} \Pr(\{Z_1 \leq \cdots \leq t\} \cap \{t \leq \cdots \leq Z_n\}) \, f_{\frac{1}{\sqrt{2}} Z}(t) \mathrm{d}t \\
&= \Pr\left(Z_1 \leq \cdots \leq Z_{i-1} \leq \frac{1}{\sqrt{2}} Z_i \leq Z_{i+2} \leq \cdots \leq Z_n \right) \\
&\overset{(d)}{=} \Pr\left(A_i \mathbf{Z} \in \mathcal{H}_{(1,2,\ldots,n-1)} \right) \\
&= \Pr\left(\mathbf{Z} \in A_i^{-1} \in \mathcal{H}_{(1,2,\ldots,n-1)} \right) \\
&\overset{(e)}{=} \frac{\mathrm{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1) \cap A_i^{-1} \mathcal{H}_{(1,2,\ldots,n-1)} \right)}{\mathrm{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1) \right)} \\
&\overset{(f)}{=} \left| \det\left(A_i^{-1} \right) \right| \frac{\mathrm{Vol}\left(A_i \mathcal{B}(\mathbf{0}_{n-1}, 1) \cap \mathcal{H}_{(1,2,\ldots,n-1)} \right)}{\mathrm{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1) \right)} \\
&\overset{(g)}{=} \sqrt{2} \frac{\mathrm{Vol}\left(\mathcal{E}(\mathbf{0}_{n-1}, i) \cap \mathcal{H}_{(1,2,\ldots,n-1)} \right)}{\mathrm{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1) \right)},
\end{aligned}
\tag{B.39}
$$

where the labeled equalities follow from: (a) the definition of $\mathbf{V}$ and writing it in terms of the standard normal random vector; (b) the law of total expectation, where we abbreviated $\{\cdots \leq Z_i\} \cap \{Z_{i+1} \leq \cdots\} \triangleq \{Z_1 \leq \cdots \leq Z_i\} \cap \{Z_{i+1} \leq \cdots \leq Z_n\}$; (c) using the fact that

$$
\begin{aligned}
f_{Z_i, Z_{i+1} | Z_i = Z_{i+1}}(t,t) &= \frac{f_{Z_i, Z_{i+1}}(t,t)}{\int_{-\infty}^{\infty} f_{Z_i, Z_{i+1}}(z,z) \, \mathrm{d}z} \\
&= f_{\frac{1}{\sqrt{2}} Z}(t);
\end{aligned}
$$

(d) letting $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_{n-1}, I_{n-1})$, defining a diagonal matrix $A_i \in \mathbb{R}^{(n-1)\times(n-1)}$ with the i-th element equal to $\frac{1}{\sqrt{2}}$ and the others equal to one, and recalling that from (2.1) we have $\mathcal{H}_{(1,2,\ldots,n-1)} = \{\mathbf{x} \in \mathbb{R}^{n-1} : x_1 \leq \cdots \leq x_{n-1}\}$; (e) using the $(n-1)$-dimensional volume

expression for the probability of a standard normal vector [1]; (f) the fact that $\mathrm{Vol}(A\mathcal{S}) = |\det(A)|\mathrm{Vol}(\mathcal{S})$ for any invertible matrix A and any set $\mathcal{S}$; and (g) letting $\mathcal{E}(\mathbf{0}_{n-1}, i)$ be 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.

Substituting (B.39) into (B.35), and noting that $f_{V_i}(0) = \frac{1}{2\sqrt{\pi}}$ for all $i \in [1 : n-1]$, we obtain

$$\lim_{\sigma\to\infty} \frac{P_{e,\mathrm{lin}}(\infty) - P_{e,\mathrm{lin}}(\sigma)}{\frac{1}{\sigma}} = \frac{1}{\sqrt{2\pi}} \sum_{\tau\in\mathcal{P}} P_{\mathbf{X}}(\mathcal{H}_\tau) \sum_{i=1}^{n-1} \alpha_i \mathbb{E}\left[(\mathbf{W}_\tau)_i\right],$$

where, for all $i \in [1 : n-1]$, we have that

$$\alpha_i = \frac{\mathrm{Vol}\left(\mathcal{E}(\mathbf{0}_{n-1}, i) \cap \mathcal{H}_{(1,2,\ldots,n-1)}\right)}{\mathrm{Vol}\left(\mathcal{B}(\mathbf{0}_{n-1}, 1)\right)}. \tag{B.40}$$

This concludes the proof of Theorem 4.5.1.

B.5 Proof of Theorem 4.6.1

Without loss of generality, we assume that $\tau = (1, 2, \ldots, n)$. Then, from Lemma 4.3.5 we can write the probability of correctness as

$$\begin{aligned} P_c(\sigma) = P_{c,\mathrm{lin}}(\sigma) &= \Pr\left(\mathbf{X} + \mathbf{N} \in \mathcal{H}_\tau \mid \mathbf{X} \in \mathcal{H}_\tau\right) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} \{X_i + N_i \le X_{i+1} + N_{i+1}\} \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} \{N_i - N_{i+1} \le X_{i+1} - X_i\} \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau\right) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} \{V_i \le W_i\}\right), \end{aligned} \tag{B.41}$$

where $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_{n-1}, \sigma^2\tilde{K})$ and $\tilde{K}$ is defined in (4.22).

We start by proving the upper bound. Since $\mathbf{V}$ is negatively associated (see Lemma B.2.1

in Appendix B.2), we have that

$$\begin{aligned} P_{c,\mathrm{lin}}(\sigma) &\overset{(a)}{\leq} \prod_{i=1}^{n-1} \Pr(V_i \leq W_i) \\ &= \prod_{i=1}^{n-1} \Pr\left(Z \leq \frac{W_i}{\sqrt{2}\sigma}\right) \\ &\overset{(b)}{\leq} \prod_{i=1}^{n-1} \Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right), \end{aligned} \tag{B.42}$$

where the labeled inequalities follow from: (a) Lemma B.2.1; and (b) using Jensen's inequality with the facts that $\Phi(x)$ is concave on $x \in \mathbb{R}_+$ and that $W_i > 0$ for all $i \in [1 : n-1]$.

We now prove the lower bound. From Lemma 4.3.5, the probability of correctness can be written as

$$\begin{aligned} P_{c,\mathrm{lin}}(\sigma) &= \Pr(\mathbf{X} + \mathbf{N} \in \mathcal{H}_\tau \mid \mathbf{X} \in \mathcal{H}_\tau) \\ &= \frac{1}{n!} + \frac{1}{n!} \sum_{\eta \in \mathcal{P} \setminus \tau} \Pr(\mathbf{X} + \mathbf{N} \in \mathcal{H}_\tau \mid \mathbf{X} \in \mathcal{H}_\tau, \mathbf{N} \in \mathcal{H}_\eta) \\ &= \frac{1}{n!} + \frac{1}{n!} \sum_{\eta \in \mathcal{P} \setminus \tau} \Pr\left(\bigcap_{i=1}^{n-1} \{N_i - N_{i+1} \leq W_i\} \,\middle|\, \mathbf{N} \in \mathcal{H}_\eta \right), \end{aligned} \tag{B.43}$$

where the last equality follows because for any $\eta \in \mathcal{P} \setminus \tau$,

$$\begin{aligned} &\Pr(\mathbf{X} + \mathbf{N} \in \mathcal{H}_\tau \mid \mathbf{X} \in \mathcal{H}_\tau, \mathbf{N} \in \mathcal{H}_\eta) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} \{X_i + N_i \leq X_{i+1} + N_{i+1}\} \,\middle|\, \mathbf{X} \in \mathcal{H}_\tau, \mathbf{N} \in \mathcal{H}_\eta \right) \\ &= \Pr\left(\bigcap_{i=1}^{n-1} \{N_i - N_{i+1} \leq W_i\} \,\middle|\, \mathbf{N} \in \mathcal{H}_\eta \right). \end{aligned} \tag{B.44}$$

We now let $Z_i \sim \mathcal{N}(0,1)$, for any $i \in [1:n-1]$, and we observe that

$$\begin{aligned} N_i - N_{i+1} &\leq |N_i| + |N_{i+1}| \\ &= \sigma(|Z_i| + |Z_{i+1}|) \\ &\leq \sigma\sqrt{2}\sqrt{Z_i^2 + Z_{i+1}^2} \\ &\leq \sigma\sqrt{2}\|\mathbf{Z}\|_2, \end{aligned} \tag{B.45}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. Then, P_c in (B.43) can be lower bounded as

$$\begin{aligned} P_{c,\text{lin}}(\sigma) &\geq \frac{1}{n!} + \frac{1}{n!}\sum_{\eta \in \mathcal{P}\setminus\tau} \Pr\left(\bigcap_{i=1}^{n-1}\{\sqrt{2}\sigma\|\mathbf{Z}\|_2 \leq W_i\} \,\middle|\, \mathbf{N} \in \mathcal{H}_\eta\right) \\ &= \frac{1}{n!} + \frac{1}{n!}\sum_{\eta \in \mathcal{P}\setminus\tau} \Pr\left(\sqrt{2}\sigma\|\mathbf{Z}\|_2 \leq \min_i\{W_i\}\right) \\ &= \frac{1}{n!} + \frac{n!-1}{n!}\Pr\left(\|\mathbf{Z}\|_2 \leq \frac{\min_i\{W_i\}}{\sqrt{2}\sigma}\right). \end{aligned} \tag{B.46}$$

Using the fact that $\|\mathbf{Z}\|_2 \stackrel{d}{=} \mathcal{X}_n$, where $\mathcal{X}_n$ is the chi distributed random variable with n degrees of freedom, we obtain the lower bound. This concludes the proof of Theorem 4.6.1.

B.6 Proof of Corollary 4.6.3

By letting $R_n \stackrel{d}{=} \sum_{i=1}^{n-1} W_i$, the upper bound in Theorem 4.6.1 can be further bounded as

$$\begin{aligned} P_c(\sigma) &\leq \prod_{i=1}^{n-1} \Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right) \\ &= \prod_{i=1}^{n-1} \mathrm{e}^{\log \Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right)} \\ &= \mathrm{e}^{\frac{n-1}{n-1}\sum_{i=1}^{n-1} \log \Phi\left(\frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right)} \\ &\stackrel{(a)}{\leq} \mathrm{e}^{(n-1)\log \Phi\left(\frac{1}{n-1}\sum_{i=1}^{n-1} \frac{\mathbb{E}[W_i]}{\sqrt{2}\sigma}\right)} \\ &= \mathrm{e}^{(n-1)\log \Phi\left(\frac{1}{n-1}\frac{\mathbb{E}[R]}{\sqrt{2}\sigma}\right)} \\ &\stackrel{(b)}{\leq} \mathrm{e}^{(n-1)\log \Phi\left(\frac{1}{n-1}\frac{2\sqrt{2\gamma^2\log(n)}}{\sqrt{2}\sigma}\right)}, \end{aligned} \tag{B.47}$$

where the labeled inequalities follow from: (a) using Jensen's inequality; and (b) the fact that for a γ^2-sub-Gaussian random variable, $\mathbb{E}[R_n] \leq 2\sqrt{2\gamma^2\log(n)}$ [62]. Then, by taking the Taylor series for $\Phi(z)$, we obtain

$$\Phi(z) \leq \frac{1}{2} + \frac{1}{\sqrt{2\pi}} z, \tag{B.48}$$

for any $z \geq 0$. Thus,

$$\begin{aligned} P_c(\sigma) &\leq \mathrm{e}^{(n-1)\log\left(\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\frac{1}{n-1}\frac{2\sqrt{2\gamma^2\log(n)}}{\sqrt{2}\sigma}\right)} \\ &= \left(\frac{1}{2} + \frac{\sqrt{2\gamma^2\log(n)}}{\sqrt{\pi}\sigma(n-1)}\right)^{n-1} \\ &= 2^{-n+1}\left(1 + \frac{2\sqrt{2\gamma^2\log(n)}}{\sqrt{\pi}\sigma(n-1)}\right)^{\frac{n-1}{\sqrt{\log(n)}}\sqrt{\log(n)}} \\ &\leq 2^{-n+1}\mathrm{e}^{\frac{2\sqrt{2}\gamma}{\sqrt{\pi}\sigma}\sqrt{\log(n)}}, \end{aligned} \tag{B.49}$$

where the last inequality follows by Lemma B.6.1 below, the proof of which can be found in Appendix B.8.4. This concludes the proof of Corollary 4.6.3.

Lemma B.6.1. *Let $x \geq 0$. Then, for any $n \geq 0$ the following bound holds*

$$\mathrm{e}^x \geq \left(1 + \frac{x}{n}\right)^n. \tag{B.50}$$

B.7 Proof of Proposition 4.6.4

We start by deriving the lower bound on $P_c(\sigma)$. From Lemma 4.3.5, we have

$$\begin{aligned} P_c(\sigma) &= \sum_{\tau \in \mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{X}+\mathbf{N})^{\mathsf{T}} \in \mathcal{H}_\tau \times \mathcal{H}_\tau\right) \\ &\geq \sum_{\tau \in \mathcal{P}} \Pr\left((\mathbf{X}, \mathbf{N})^{\mathsf{T}} \in \mathcal{H}_\tau \times \mathcal{H}_\tau\right) \\ &= n! \frac{1}{(n!)^2}, \end{aligned} \tag{B.51}$$

where the inequality follows since, if $(\mathbf{X}, \mathbf{N})^{\mathsf{T}} \in \mathcal{H}_\tau^2$, then $(\mathbf{X}, \mathbf{X}+\mathbf{N})^{\mathsf{T}} \in \mathcal{H}_\tau^2$, where $\mathcal{H}_\tau^2 = \mathcal{H}_\tau \times \mathcal{H}_\tau$. Now to show the upper bound we use [1, Theorem 2]. We have

$$\begin{aligned} P_c(\sigma) &= n! \frac{\mathrm{Vol}\left(\mathcal{H}_\tau^2 \cap A\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)}{\sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)} \\ &\overset{(a)}{\leq} n! \frac{\mathrm{Vol}\left(\mathcal{H}_\tau^2 \cap \mathcal{B}^{2n}(\mathbf{0}_{2n}, \|A\|)\right)}{\sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)} \\ &\overset{(b)}{=} \frac{\mathrm{Vol}\left(\mathcal{B}^{2n}(\mathbf{0}_{2n}, \|A\|)\right)}{n! \sigma^n \mathrm{Vol}\left(\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right)\right)} \\ &\overset{(c)}{=} \frac{\|A\|^{2n}}{n! \sigma^n}, \end{aligned} \tag{B.52}$$

where the labeled (in)equalities follow from: (a) letting $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & \sigma I_n \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$ and the fact that $A\mathcal{B}^{2n}\left(\mathbf{0}_{2n}, 1\right) \subseteq \mathcal{B}^{2n}\left(\mathbf{0}_{2n}, \|A\|\right)$; (b) computing the volume of the intersection of a ball and a cone $\mathcal{H}_\tau$; 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

$$\|A\| = \left(\frac{(\sigma^4 + 4)^{\frac{1}{2}}}{2} + \frac{\sigma^2}{2} + 1 \right)^{\frac{1}{2}}. \tag{B.53}$$

This concludes the proof of Proposition 4.6.4.

B.8 Proof of Auxiliary Results

B.8.1 Examples for the High-Noise Regime

From [60], for $X_i \sim \text{Unif}(a, b),\ 0 \leq a < b < \infty$, we have that

$$\mathbb{E}[W_i] = \frac{b-a}{n+1}. \tag{B.54}$$

The spacings of an i.i.d. exponential random variable with parameter λ are exponential random variables with parameters $\lambda(n-1), \lambda(n-2), \ldots, \lambda$ [60]. Thus, for $X_i \sim \text{Exp}(\lambda)$,

$$\mathbb{E}[W_i] = \frac{1}{\lambda(n-i)}. \tag{B.55}$$

The key of the proof of the range R_n is to use the following expressions from [53],

$$\mathbb{E}[X_{1:n}] = n \int_{-\infty}^{\infty} x(1 - F(x))^{n-1} f(x)\, \mathrm{d}x, \tag{B.56}$$

and

$$\mathbb{E}[X_{n:n}] = n \int_{-\infty}^{\infty} x F(x)^{n-1} f(x)\, \mathrm{d}x. \tag{B.57}$$

First, consider $X_i \sim \text{Unif}(a, b),\ 0 \leq a < b < \infty$. Then,

$$\mathbb{E}[X_{1:n}] = \frac{b + an}{n+1} \text{ and } \mathbb{E}[X_{n:n}] = \frac{a + bn}{n+1}, \tag{B.58}$$

and hence, we obtain

$$\mathbb{E}[R_n] = \frac{(b-a)(n-1)}{n+1}. \tag{B.59}$$

Next, let $X_i \sim \text{Exp}(\lambda)$. Then,

$$\mathbb{E}[X_{1:n}] = \frac{1}{\lambda n} \text{ and } \mathbb{E}[X_{n:n}] = \sum_{k=1}^{n} \frac{1}{\lambda k}, \tag{B.60}$$

and hence, we obtain

$$\mathbb{E}[R_n] = \sum_{k=1}^{n-1} \frac{1}{\lambda k}. \tag{B.61}$$

B.8.2 Proof of Lemma B.3.1

The joint density function $f_{W_i,W_j}(u,v),\ 1 \leq i < j-1 \leq n-1$ is given by [60]

$$\begin{aligned} f_{W_i,W_j}(u,v) = n! \int_{-\infty}^{\infty} \int_{x+u}^{\infty} & \frac{F_X(x)^{i-1}}{(i-1)!} \frac{(F_X(y) - F_X(x+u))^{j-i-2}}{(j-i-2)!} \\ & \times \frac{(1-F_X(y+v))^{n-j-1}}{(n-j-1)!} \\ & \times f_X(x) f_X(x+u) f_X(y) f_X(y+v)\, \mathrm{d}y\, \mathrm{d}x, \end{aligned} \tag{B.62}$$

and for $i = j-1$, we have that [60]

$$\begin{aligned} f_{W_i,W_j}(u,v) = n! \int_{-\infty}^{\infty} & \frac{F_X(x)^{i-1}}{(i-1)!} \frac{(1-F_X(x+u+v))^{n-i-2}}{(n-i-2)!} \\ & \times f_X(x) f_X(x+u) f_X(x+u+v)\, \mathrm{d}x, \end{aligned} \tag{B.63}$$

where $F_X(\cdot)$ is the cdf of X and $f_X(\cdot)$ is the pdf of X. By using the upper bounds $F_X(x) \leq 1$, $1 - F_X(x) \leq 1$, and $F_X(y) - F_X(x+u) \leq 1$ for $y \geq x+u$, we obtain that, for $i < j-1$,

$$\begin{aligned} f_{W_i,W_j}(u,v) &\leq n! \int_{-\infty}^{\infty} \int_{x+u}^{\infty} \frac{f_X(x) f_X(x+u) f_X(y) f_X(y+v)}{(i-1)!(j-i-2)!(n-j-1)!} \, \mathrm{d}y \, \mathrm{d}x \\ &\leq n! \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f_X(x) f_X(x+u) f_X(y) f_X(y+v)}{(i-1)!(j-i-2)!(n-j-1)!} \, \mathrm{d}y \, \mathrm{d}x \\ &\leq \frac{n! \|f_X\|_2^4}{(i-1)!(j-i-2)!(n-j-1)!} \\ &< \infty, \end{aligned} \tag{B.64}$$

where the second inequality follows since the integrand is always non-negative, the third inequality follows from the Cauchy–Schwarz inequality, and the last inequality is due to the assumption that $\|f_X\|_2 < \infty$. Similarly, for $i = j-1$, we have that

$$\begin{aligned} f_{W_i,W_j}(u,v) &\leq n! \int_{-\infty}^{\infty} \frac{f_X(x) f_X(x+u) f_X(x+u+v)}{(i-1)!(n-i-2)!} \, \mathrm{d}x \\ &\leq \frac{n! \|f_X\|_2^3}{(i-1)!(n-i-2)!} \\ &< \infty, \end{aligned} \tag{B.65}$$

where the second inequality follows from the Cauchy–Schwarz inequality, and the last inequality follows since $\|f_X\|_2 < \infty$. This shows that the joint density is bounded everywhere.

For the marginal density, we take any $(i,j) \in [1:n]^2$ such that $i < j-1$. Then, we have that

$$\begin{aligned} f_{W_i}(u) &= \int_{-\infty}^{\infty} f_{W_i,W_j}(u,v) \, \mathrm{d}v \\ &\leq n! \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f_X(x) f_X(x+u) f_X(y) f_X(y+v)}{(i-1)!(j-i-2)!(n-j-1)!} \mathrm{d}y \mathrm{d}x \mathrm{d}v \\ &= \frac{n! \int_{-\infty}^{\infty} f_X(x) f_X(x+u) \, \mathrm{d}x}{(i-1)!(j-i-2)!(n-j-1)!} \\ &\leq \frac{n! \|f_X\|_2^2}{(i-1)!(j-i-2)!(n-j-1)!} \\ &< \infty, \end{aligned} \tag{B.66}$$

where the first inequality follows from the fact that we have shown above that

$$f_{W_i,W_j}(u,v) \leq n! \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f_X(x)f_X(x+u)f_X(y)f_X(y+v)}{(i-1)!(j-i-2)!(n-j-1)!}\,\mathrm{d}y\,\mathrm{d}x,$$

the second inequality is due to the Cauchy–Schwarz inequality, and the last inequality follows from the assumption $\|f_X\|_2 < \infty$. This concludes the proof of Lemma B.3.1.

B.8.3 Proof of Lemma B.4.1

The proof is based on Leibniz's integral rule. Let us define a function

$$g_t(\kappa,\mathbf{x}) = \int_{c_t\kappa}^{a} \cdots \int_{c_n\kappa}^{a} f(\mathbf{x})\,\mathrm{d}x_n \cdots\,\mathrm{d}x_t,\ t \in [1:n]. \tag{B.67}$$

Then, we observe that for $t = 1$, we have

$$\begin{aligned}\frac{\partial}{\partial\kappa} g_1(\kappa,\mathbf{x}) &= \frac{\partial}{\partial\kappa}\int_{c_1\kappa}^{a} g_2(\kappa,\mathbf{x})\,\mathrm{d}x_1 \\ &= \int_{c_1\kappa}^{a} \frac{\partial}{\partial\kappa} g_2(\kappa,\mathbf{x})\,\mathrm{d}x_1 - c_1 g_2(\kappa,\mathbf{x})|_{x_1=c_1\kappa},\end{aligned} \tag{B.68}$$

where the last equality follows by using the Leibniz's integral rule. Moreover, for any $t \in [1 : n-1]$, the Leibniz's integral rule indicates that

$$\begin{aligned}\frac{\partial}{\partial\kappa} g_t(\kappa,\mathbf{x}) &= \frac{\partial}{\partial\kappa}\int_{c_t\kappa}^{a} g_{t+1}(\kappa,\mathbf{x})\,\mathrm{d}x_t \\ &= \int_{c_t\kappa}^{a} \frac{\partial}{\partial\kappa} g_{t+1}(\kappa,\mathbf{x})\,\mathrm{d}x_t - c_t g_{t+1}(\kappa,\mathbf{x})|_{x_t=c_t\kappa},\end{aligned} \tag{B.69}$$

and for $t = n$, by the fundamental theorem of calculus, we have

$$\begin{aligned}\frac{\partial}{\partial\kappa} g_n(\kappa,\mathbf{x}) &= \frac{\partial}{\partial\kappa}\int_{c_n\kappa}^{a} f(\mathbf{x})\,\mathrm{d}x_n \\ &= -c_n f(\mathbf{x})|_{x_n=c_n\kappa}.\end{aligned} \tag{B.70}$$

Hence, by mathematical induction using (B.68), (B.69), and (B.70), we get

$$\begin{aligned}
\frac{\partial}{\partial\kappa}g_1(\kappa,\mathbf{x}) &= \int_{c_1\kappa}^{a}\frac{\partial}{\partial\kappa}g_2(\kappa,\mathbf{x})\,\mathrm{d}x_1 - c_1 g_2(\kappa,\mathbf{x})|_{x_1=c_1\kappa} \\
&= \int_{c_1\kappa}^{a}\left(\int_{c_2\kappa}^{a}\frac{\partial}{\partial\kappa}g_3(\kappa,\mathbf{x})\,\mathrm{d}x_2 - c_2 g_3(\kappa,\mathbf{x})|_{x_2=c_2\kappa}\right)\mathrm{d}x_1 \\
&\quad - c_1 g_2(\kappa,\mathbf{x})|_{x_1=c_1\kappa} \\
&= \int_{c_1\kappa}^{a}\int_{c_2\kappa}^{a}\frac{\partial}{\partial\kappa}g_3(\kappa,\mathbf{x})\,\mathrm{d}x_2\,\mathrm{d}x_1 \\
&\quad - c_2\int_{c_1\kappa}^{a} g_3(\kappa,\mathbf{x})|_{x_2=c_2\kappa}\,\mathrm{d}x_1 - c_1 g_2(\kappa,\mathbf{x})|_{x_1=c_1\kappa} \\
&\quad \vdots \\
&= -\sum_{i=1}^{n} c_i\int_{c_1\kappa}^{a}\cdots\int_{c_n\kappa}^{a} f(\mathbf{x})|_{x_i=c_i\kappa}\mathrm{d}x_n\cdots\mathrm{d}x_{i+1}\mathrm{d}x_{i-1}\cdots\mathrm{d}x_1. \qquad \text{(B.71)}
\end{aligned}$$

This concludes the proof of Lemma B.4.1.

B.8.4 Proof of Lemma B.6.1

Since $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \mathrm{e}^x$, it is sufficient to show that $\left(1+\frac{x}{n}\right)^n$ is a non-decreasing function in n when $x \geq 0$. Taking the derivative with respect to n, we obtain

$$\frac{\partial}{\partial n}\left(1+\frac{x}{n}\right)^n = \left(1+\frac{x}{n}\right)^n\left(\log\left(1+\frac{x}{n}\right) - \frac{x}{x+n}\right). \qquad \text{(B.72)}$$

Since the first term in (B.72) is positive, we only need to verify that the second term is positive. In other words, we need to prove the following,

$$\log\left(1+\frac{x}{n}\right) - \frac{x}{x+n} \overset{?}{\geq} 0,\ \forall x \geq 0, n \geq 0. \qquad \text{(B.73)}$$

This can be easily verified by using the concavity property of $\log(x)$. In particular, we have the first-order condition for the concave function $\log(u)$ as

$$\begin{aligned}
&\log(v) \leq \log(u) + \frac{1}{u}(v-u) \\
\Longrightarrow\ &\log\left(\frac{u}{v}\right) \geq \frac{u-v}{u},\ \forall u, v > 0. \qquad \text{(B.74)}
\end{aligned}$$

Taking $u = x + n$ and $v = n$, we get

$$\log\left(\frac{x+n}{n}\right) \geq \frac{x}{x+n}, \tag{B.75}$$

which verifies (B.73). This concludes the proof of Lemma B.6.1.

Appendix C

Differed Proofs in Chapter 5

C.1 Proof of Example 5.2.5

C.1.1 Gaussian Noise

If $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$, its PDF is

$$f_{\mathbf{N}}(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}} = g(\|\mathbf{z}\|_2), \tag{C.1}$$

with $g(t) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{t^2}{2\sigma^2}}$. Since $g(t)$ is a non-increasing function in $t > 0$, then $\mathcal{N}(\mathbf{0}_n, \sigma^2 I_n) \in \mathcal{S}_{n,2}$.

C.1.2 Laplace Noise

If $\mathbf{N}$ consists of i.i.d. $N_i \sim \text{Lap}(0, b)$, its PDF is

$$f_{\mathbf{N}}(\mathbf{z}) = \prod_{i=1}^{n} \frac{1}{2b} e^{-\frac{|z_i|}{b}} = \frac{1}{(2b)^n} e^{-\frac{\|\mathbf{z}\|_1}{b}} = g(\|\mathbf{z}\|_1), \tag{C.2}$$

with $g(t) = \frac{1}{(2b)^n} e^{-\frac{t}{b}}$, which is a non-increasing function in $t > 0$. Thus, a joint distribution of i.i.d. $\text{Lap}(0, b)$ is a member of $\mathcal{S}_{n,1}$.

C.1.3 Generalized Normal Noise

If $\mathbf{N}$ consists of i.i.d. $N_i \sim \mathcal{GN}(0, a, p)$, its PDF is [75]

$$f_{\mathbf{N}}(\mathbf{z}) = \prod_{i=1}^{n} K \exp\left(-\left|\frac{z_i}{a}\right|^p\right) = K^n \exp\left(-\frac{\sum_{i=1}^{n} |z_i|^p}{a^p}\right) = g(\|\mathbf{z}\|_p), \tag{C.3}$$

where $K = \frac{p}{2a\Gamma(1/p)}$ is the normalization factor, and $g(t) = K^n \exp\left(-\frac{t^p}{a^p}\right)$, which is a non-increasing function in $t > 0$. Thus, a joint distribution of i.i.d. $\mathcal{GN}(0, a, p)$ is a member of $\mathcal{S}_{n,p}$.

C.1.4 Staircase Noise

If $\mathbf{N}$ has a staircase distribution with $(\lambda, \gamma, \Delta)$, its PDF is of the form [72]

$$f_{\mathbf{N}}(\mathbf{z}) = \beta e^{-\lambda h(\mathbf{z})}, \tag{C.4}$$

with

$$h(\mathbf{z}) = \begin{cases} k & \text{if } \|\mathbf{z}\|_1 \in [k\Delta, (k+\gamma)\Delta], \\ k+1 & \text{if } \|\mathbf{z}\|_1 \in [(k+\gamma)\Delta, (k+1)\Delta], \end{cases} \tag{C.5}$$

where $\gamma \in [0, 1]$ and $\Delta > 0$ are given, and β is a normalization parameter.

Since $h(\mathbf{z})$ is a non-decreasing function in $\|\mathbf{z}\|_1$, then $f_{\mathbf{N}}(\mathbf{z})$ is non-increasing in $\|\mathbf{z}\|_1$. Thus, a staircase distribution is a member of $\mathcal{S}_{n,1}$.

C.1.5 Uniform Noise

If $\mathbf{N} \sim \text{Unif}(\mathcal{B}_p(\mathbf{0}_n, r))$ with $r > 0$, its PDF is given by

$$\begin{aligned} f_{\mathbf{N}}(\mathbf{z}) &= \frac{1}{\text{Vol}(\mathcal{B}_p(\mathbf{0}_n, r))} \mathbb{1}_{\{\mathbf{z} \in \mathcal{B}_p(\mathbf{0}_n, r)\}} \\ &= \frac{1}{\text{Vol}(\mathcal{B}_p(\mathbf{0}_n, r))} \mathbb{1}_{\{\|\mathbf{z}\|_p \le r\}}, \end{aligned} \tag{C.6}$$

where $\text{Vol}(\mathcal{S})$ denotes the volume of the set $\mathcal{S}$. Clearly, the PDF in (C.6) is a non-increasing function in $\|\mathbf{z}\|_p$ and hence, $\text{Unif}(\mathcal{B}_p(\mathbf{0}_n, r))$ is $\mathcal{S}_{n,p}$.

C.2 Proof of Lemma 5.3.2

Our goal is to show that a solution $\hat{\kappa}$ for the following optimization problem,

$$\arg\min_{\kappa\in\mathcal{P}} \|\mathbf{y} - P_{\eta\to\kappa}\mathbf{x}\|_p, \tag{C.7}$$

for any $\eta \in \mathcal{P}$, $\mathbf{y} \in \mathcal{H}_\tau$ and $\mathbf{x} \in \mathcal{H}_\eta$, is given by $\hat{\kappa} = \tau$. We start by noting that, by the property of permutation invariance of the ℓ_p-norm, without loss of generality, we can consider $\tau = (1, 2, \ldots, n)$ which indicates that $\mathbf{y}$ is sorted in ascending order. In addition, a solution to (C.7) does not depend on the permutation of $\mathbf{x}$, i.e., we can start the problem with any $\mathbf{x} \in \mathcal{H}_\eta$. This is beause we consider every possible permutation matrix $P_{\eta\to\kappa}$'s with $\kappa \in \mathcal{P}$. Hence, we set $\eta = \tau$, which implies that $\mathbf{x}$ is also sorted according to the ascending order, i.e., $\mathbf{x}$ and $\mathbf{y}$ are sorted according to the same permutation τ.

The key tool that will enable our proof is the following generalized version of the rearrangement inequality [86].

Lemma C.2.1. *Consider a sequence of real numbers $a_1 \leq \ldots \leq a_n$ and a collection of functions $f_i(\cdot) : [a_1, a_n] \mapsto \mathbb{R}$ for $i \in [1 : n]$ and for some fixed n. Suppose that for all $x \in [a_1, a_n]$ we have that*

$$f_1'(x) \leq \ldots \leq f_n'(x). \tag{C.8}$$

Then, for any permutation $\kappa \in \mathcal{P}$, we have that

$$\sum_{i=1}^{n} f_i(a_{n-i+1}) \leq \sum_{i=1}^{n} f_i(a_{\kappa_{n-i+1}}), \tag{C.9}$$

where $a_{\kappa_i} = (P_{\tau\to\kappa}\mathbf{a})_i$, $i \in [1 : n]$.

For the given $\mathbf{y} \in \mathcal{H}_\tau$, in order to apply Lemma C.2.1, we define a sequence of functions

$$f_i(t) \triangleq |y_{n-i+1} - t|^p, \ i \in [1 : n]. \tag{C.10}$$

Note that $y_1 \leq y_2 \leq \cdots \leq y_n$. For the time being assume that

$$f_i'(t) \leq f_j'(t), \ \forall t \in \mathbb{R}, \tag{C.11}$$

for all $i < j$. The claim in (C.11), which will be shown later, guarantees that we can use Lemma C.2.1. Therefore, by setting $a_i = x_i$ in Lemma C.2.1, and recalling that $\mathbf{x}$ and $\mathbf{y}$ are sorted according to the same permutation τ, we arrive at

$$\begin{aligned}
\|\mathbf{y} - \mathbf{x}\|_p^p &= \sum_{i=1}^{n} |y_{n-i+1} - x_{n-i+1}|^p \\
&= \sum_{i=1}^{n} f_i(x_{n-i+1}) \\
&\leq \sum_{i=1}^{n} f_i(x_{\kappa_{n-i+1}}) \\
&= \sum_{i=1}^{n} |y_{n-i+1} - x_{\kappa_{n-i+1}}|^p \\
&= \sum_{i=1}^{n} |y_{n-i+1} - (P_{\tau\to\kappa}\mathbf{x})_{n-i+1}|^p \\
&= \|\mathbf{y} - P_{\tau\to\kappa}\mathbf{x}\|_p^p,
\end{aligned}$$

for all $\kappa \in \mathcal{P}$. This indeed shows that, under the assumption in (C.11), a solution $\hat{\kappa}$ for the optimization problem in (C.7) is given by $\hat{\kappa} = \tau$.

To complete the proof it remains to verify that the condition in (C.11) holds. Towards this end, we observe that

$$f_i'(t) = p(t - y_{n-i+1})|t - y_{n-i+1}|^{p-2},$$

for all $i \in [1 : n]$. We now show that $f_i'(t) \leq f_j'(t)$ for all $t \in \mathbb{R}$ and $i < j$. This follows by a simple comparison, which consists of subtracting $f_j'(t)$ from $f_i'(t)$, namely

$$f_i'(t) - f_j'(t) = p(t - y_{n-i+1})|t - y_{n-i+1}|^{p-2} - p(t - y_{n-j+1})|t - y_{n-j+1}|^{p-2}, \quad \text{(C.12)}$$

where $y_{n-i+1} \geq y_{n-j+1}$ since $\mathbf{y}$ by assumption is sorted in ascending order. If (C.12) is less than or equal to zero, then (C.11) holds. We now show that (C.12) is indeed always less than or equal to zero.

- $t \in [-\infty, y_{n-j+1}]$: In this case, (C.12) becomes

$$f_i'(t) - f_j'(t) = -p\left(y_{n-i+1} - t\right)^{p-1} + p\left(y_{n-j+1} - t\right)^{p-1},$$

  which is always less than or equal to zero;

- $t \in [y_{n-j+1}, y_{n-i+1}]$: In this case, (C.12) becomes

$$f_i'(t) - f_j'(t) = -p\left(y_{n-i+1} - t\right)^{p-1} - p\left(t - y_{n-j+1}\right)^{p-1},$$

  which is always less than or equal to zero;

- $t \in [y_{n-i+1}, \infty]$: In this case, (C.12) becomes

$$f_i'(t) - f_j'(t) = -p\left(t - y_{n-i+1}\right)^{p-1} - p\left(t - y_{n-j+1}\right)^{p-1},$$

  which is always less than or equal to zero.

The above three cases imply that the inequality in (C.11) holds for any $i < j$ and hence, the sequence of functions in (C.10) satisfies (C.8). This concludes the proof of the desired claim and the proof of Lemma 5.3.2.

C.3 Proof of Theorem 5.3.6

The probability of error is given by [3, Corollary 1],

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = 1 - \mathbb{E}\left[\Pr\left(\mathbf{V} \geq -\frac{T_\tau \mathbf{X}}{\sigma} \;\middle|\; \mathbf{X}\right) \;\middle|\; \mathbf{X} \in \mathcal{H}_\tau\right], \tag{C.13}$$

where T_τ is defined in (5.15). We now note that, from Remark 5.3.5, we can write $T_\tau \mathbf{X}|\mathbf{X} \in \mathcal{H}_\tau$ as a spacing vector $\mathbf{W}$ and hence, we can equivalently rewrite (C.13) as

$$\begin{aligned} P_e(\phi_{\text{lin}}, \mathcal{K}) &= 1 - \Pr\left(\bigcap_{i=1}^{n-1} \left\{ V_i \geq \frac{-W_i}{\sigma} \right\}\right) \\ &= \Pr\left(\bigcup_{i=1}^{n-1} \left\{ V_i < \frac{-W_i}{\sigma} \right\}\right) \\ &= \sum_{k=1}^{n-1} \left((-1)^{k-1} \sum_{\substack{\mathcal{I} \subseteq [1:n-1] \\ |\mathcal{I}|=k}} \Pr\left(\mathcal{A}_\mathcal{I}\right) \right), \end{aligned} \tag{C.14}$$

where the last equality follows from the inclusion-exclusion principle where $\mathcal{A}_\mathcal{I} = \cap_{i \in \mathcal{I}} \mathcal{A}_i$ with $\mathcal{A}_i = \{V_i < -\sigma^{-1} W_i\}$.

For any set $|\mathcal{I}| = k$, we have

$$\begin{aligned} \Pr(\mathcal{A}_\mathcal{I}) &= \Pr\left(\bigcap_{i \in \mathcal{I}} \left\{ V_i < \frac{-W_i}{\sigma} \right\}\right) \\ &= \int_{\mathbf{w} \in \mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}\left(\frac{-\mathbf{w}}{\sigma}\right) f_{\mathbf{W}_\mathcal{I}}(\mathbf{w}) \mathrm{d}\mathbf{w} \\ &= \int_{\mathbf{u} \in \mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}\left(-\mathbf{u}\right) f_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{u}) \sigma^k \mathrm{d}\mathbf{u}, \end{aligned} \tag{C.15}$$

where the last equality follows from a change of variable. By substituting (C.15) into (C.14), we obtain

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \sum_{k=1}^{n-1} (-1)^{k-1} \alpha_k(\sigma) \sigma^k, \tag{C.16}$$

where

$$\alpha_k(\sigma) = \sum_{\substack{\mathcal{I} \subseteq [1:n-1] \\ |\mathcal{I}|=k}} \int_{\mathbf{u} \in \mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}(-\mathbf{u}) f_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{u}) \mathrm{d}\mathbf{u}. \tag{C.17}$$

Let $f^{(m)}_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{u}) = \frac{\partial^m}{\partial \sigma^m} f_{\mathbf{W}_\mathcal{I}}(\sigma \mathbf{u})$. By the Leibniz integral rule, the m-th derivative of $\alpha_k(\sigma)$ w.r.t.

σ is

$$\alpha_k^{(m)}(\sigma) = \sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=k}} \int_{\mathbf{u}\in\mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}(-\mathbf{u}) f_{\mathbf{W}_\mathcal{I}}^{(m)}(\sigma\mathbf{u})\mathrm{d}\mathbf{u}. \tag{C.18}$$

Since (C.18) is bounded at $\sigma \to 0^+$ (i.e., $\lim_{\sigma\to 0^+} |\alpha_k^{(m)}(\sigma)| < \infty$), after some trivial algebra, we obtain that the m-th derivative of $P_e(\phi_{\mathrm{lin}}, \mathcal{K})$ in (C.16) at $\sigma \to 0^+$ is

$$P_e^{(m)}(\sigma)\Big|_{\sigma\to 0^+} = \sum_{k=1}^{\min\{m,n-1\}} (-1)^{k-1} \binom{m}{k} k! \alpha_k^{(m-k)}(0^+). \tag{C.19}$$

We conclude the proof of Theorem 5.3.6 by plugging (C.19) into the Taylor series of $P_e(\phi_{\mathrm{lin}}, \mathcal{K})$ at $\sigma = 0^+$.

C.4 Proof of Corollary 5.3.7

Since $P_e(\phi_{\mathrm{lin}}, \mathcal{K}) \to 0$ as $\sigma \to 0^+$, the first order rate is given from Theorem 5.3.6 by

$$\begin{aligned}
P_e^{(1)} &= \lim_{\sigma\to 0^+} \sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=1}} \int_{\mathbf{u}\in\mathbb{R}_+} F_{\mathbf{V}_\mathcal{I}}(-\mathbf{u}) f_{\mathbf{W}_\mathcal{I}}(\sigma\mathbf{u})\, \mathrm{d}\mathbf{u} \\
&= \lim_{\sigma\to 0^+} \sum_{i=1}^{n-1} \int_{u\in\mathbb{R}_+} \Pr(V_i \le -u) f_{W_i}(\sigma u)\, \mathrm{d}u \\
&\overset{\text{(a)}}{=} \lim_{\sigma\to 0^+} \sum_{i=1}^{n-1} \int_{v\in\mathbb{R}_+} Q(v) f_{W_i}(\sqrt{2}\sigma v)\sqrt{2}\, \mathrm{d}v \\
&\overset{\text{(b)}}{=} \sum_{i=1}^{n-1} f_{W_i}(0^+) \int_{v\in\mathbb{R}_+} Q(v)\sqrt{2}\, \mathrm{d}v \\
&= \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}},
\end{aligned} \tag{C.20}$$

where (a) follows using the change of variable $u = \sqrt{2}v$ toegther with the fact that $V_i \sim \mathcal{N}(0,2)$ with $Q(\cdot)$ being the Q function of the standard normal distribution; and (b) follows by the dominated convergence theorem.

The second order rate is also given from Theorem 5.3.6 by

$$\frac{1}{2}P_e^{(2)} = \frac{1}{2}\sum_{k=1}^{2}(-1)^{k-1}\binom{2}{k}k!\alpha_k^{(2-k)}(0^+) = \alpha_1^{(1)}(0^+) - \alpha_2(0^+). \tag{C.21}$$

We need to compute $\alpha_1^{(1)}(0^+)$ and $\alpha_2(0^+)$. Firstly, we have

$$\begin{aligned}
\alpha_1^{(1)}(0^+) &= \lim_{\sigma\to 0^+}\sum_{i=1}^{n-1}\int_0^\infty F_{V_i}(-u)f_{W_i}^{(1)}(\sigma u)\,\mathrm{d}u \\
&\overset{\text{(a)}}{=} \lim_{\sigma\to 0^+}\sum_{i=1}^{n-1}\int_{u\in\mathbb{R}_+}\Pr(V_i\le -u)uf'_{W_i}(\sigma u)\,\mathrm{d}u \\
&\overset{\text{(b)}}{=} \sum_{i=1}^{n-1}f'_{W_i}(0^+)\int_{u\in\mathbb{R}_+}\Pr(V_i\le -u)u\,\mathrm{d}u \\
&\overset{\text{(c)}}{=} \sum_{i=1}^{n-1}f'_{W_i}(0^+)\int_{v\in\mathbb{R}_+}Q(v)2v\,\mathrm{d}v \\
&= \frac{1}{2}\sum_{i=1}^{n-1}f'_{W_i}(0^+),
\end{aligned} \tag{C.22}$$

where the labeled equalities follow from: (a) letting $f'_{W_i}(\sigma u) = \frac{\partial}{\partial w}f_{W_i}(w)\big|_{w=\sigma u}$; (b) using the dominated convergence theorem; and (c) using the change of variable $u = \sqrt{2}v$ similar to the step (a) in (C.20).

The second term in (C.21) is given from Theorem 5.3.6 by

$$\begin{aligned}
\alpha_2(0^+) &= \lim_{\sigma\to 0^+} \sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=2}} \int_{\mathbf{u}\in\mathbb{R}_+^2} \Pr(\mathbf{V}_{\mathcal{I}} \le -\mathbf{u}) f_{\mathbf{W}_{\mathcal{I}}}(\sigma\mathbf{u})\,\mathrm{d}\mathbf{u} \\
&\overset{(a)}{=} \sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=2}} f_{\mathbf{W}_{\mathcal{I}}}(\mathbf{0}_2^+) \int_{\mathbf{u}\in\mathbb{R}_+^2} \Pr(\mathbf{V}_{\mathcal{I}} \le -\mathbf{u})\,\mathrm{d}\mathbf{u} \\
&= \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) \int_{\mathbf{u}\in\mathbb{R}_+^2} \Pr(V_i \le -u_1, V_{i+1} \le -u_2)\,\mathrm{d}\mathbf{u} \\
&\quad + \sum_{\substack{(i,j)\in[1:n-1]^2\\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+) \int_{\mathbf{u}\in\mathbb{R}_+^2} \Pr(V_i \le -u_1, V_j \le -u_2)\,\mathrm{d}\mathbf{u} \\
&\overset{(b)}{=} \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) \int_{\mathbf{u}\in\mathbb{R}_+^2} Q_{V_i,V_{i+1}}(\mathbf{u})\,\mathrm{d}\mathbf{u} + \sum_{\substack{(i,j)\in[1:n-1]^2\\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+) \left(\int_{u\in\mathbb{R}_+} Q_V(u)\,\mathrm{d}u\right)^2 \\
&= \mathbb{E}[\max\{0,V_1\}\max\{0,V_2\}] \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) + \mathbb{E}^2[\max\{0,V_1\}] \sum_{\substack{(i,j)\in[1:n-1]^2\\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+) \\
&\overset{(c)}{\approx} 0.108998 \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) + \frac{1}{\pi} \sum_{\substack{(i,j)\in[1:n-1]^2\\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+), \qquad \text{(C.23)}
\end{aligned}$$

where (a) follows by the dominated convergence theorem; (b) is due to the independence of V_i and V_j if $|i-j|>1$ (since $V_i = N_{i+1} - N_i$ and $V_j = N_{j+1} - N_j$ with i.i.d. $\mathbf{N}$); (c) follows by noting that $\mathbb{E}[\max\{0,V_1\}\max\{0,V_2\}] \approx 0.108998$ and $\mathbb{E}^2[\max\{0,V_1\}] = \frac{1}{\pi}$.

By substituting (C.22) and (C.23) into (C.21), we obtain

$$\frac{1}{2}P_e^{(2)} \approx \frac{1}{2}\sum_{i=1}^{n-1} f'_{W_i}(0^+) - 0.108998 \sum_{i=1}^{n-2} f_{W_i,W_{i+1}}(\mathbf{0}_2^+) - \frac{1}{\pi} \sum_{\substack{(i,j)\in[1:n-1]^2\\ j>i+1}} f_{W_i,W_j}(\mathbf{0}_2^+). \qquad \text{(C.24)}$$

This concludes the proof of Corollary 5.3.7.

C.5 Simulation Results and Proof of Example 5.3.8 and 5.3.9

C.5.1 Simulation Results shown in Figure C.1

For the simulations illustrated in Figure C.1, we set $\mathrm{Unif}(0,1)$ and $\mathrm{Exp}(1)$ for $X_i, i \in [1:n]$ and $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 I_n)$. The curves for the true error probability $P_e(\phi_{\mathrm{lin}})$ were obtained by Monte-Carlo simulation using 10^6 iterations, whereas we obtained the curves for the first and second order approximations by evaluating the expression in Corollary 5.3.7. The data dimension is set to $n = 10$ for (a) and (b), and to $n = 20$ for (c) and (d). It is shown from (a) and (c) that the first and second-order approximations well fit the true $P_e(\phi_{\mathrm{lin}})$ around $\sigma = 0$. Further, (b) and (d) show the approximations in the low-noise regime and illustrate that, if the targeted error probability is small, then the first-order approximation is very close to the true error probability.

C.5.2 Proof of Example 5.3.8

To evaluate the expression in Corollary 5.3.7, we need $f_{W_i}(0^+)$, $f'_{W_i}(0^+)$ and $f_{W_i,W_j}(\mathbf{0}_2^+)$. For $X_i \sim \mathrm{Unif}(0,1)$, the PDF of the spacing W_i and W_j, W_i are given by [60]

$$f_{W_i}(w) = n(1-w)^{n-1},\ \forall i \in [1:n-1], \tag{C.25}$$

$$f_{W_i,W_j}(u,v) = n(n-1)(1-u-v)^{n-2},\ \forall i \neq j, \tag{C.26}$$

which gives

$$\begin{aligned} f_{W_i}(0^+) &= n,\ \forall i, \\ f_{W_i,W_j}(\mathbf{0}_2^+) &= n(n-1),\ \forall i \neq j. \end{aligned}$$

By differentiating (C.25) with respect to w, we also obtain

$$\begin{aligned} f'_{W_i}(w) &= -n(n-1)(1-w)^{n-2}, \\ \implies f'_{W_i}(0^+) &= -n(n-1),\ \forall i. \end{aligned}$$

This concludes the proof of Example 5.3.8

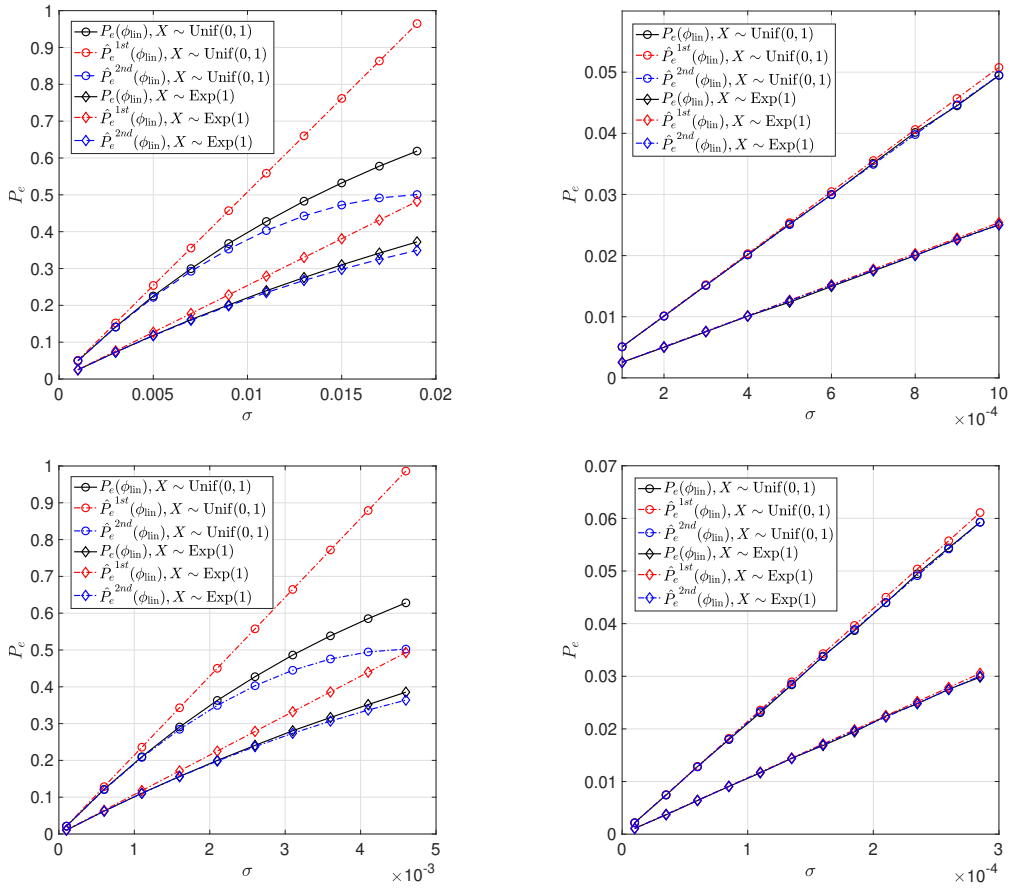


Figure C.1: Comparison between $P_e(\phi_{\text{lin}})$, the first-order approximation $\hat{P}_e^{1st}(\phi_{\text{lin}})$, and the second-order approximation $\hat{P}_e^{2nd}(\phi_{\text{lin}})$. We set $X_i \sim \text{Unif}(0,1)$ and $X_i \sim \text{Exp}(1)$ for $i \in [1:n]$: (a) $n = 10$; (b) $n = 10$ in low-noise; (c) $n = 20$; (d) $n = 20$ in low-noise.

C.5.3 Proof of Example 5.3.9

To evaluate c_1 and c_2, we make use of the fact [60] that the spacings of $\text{Exp}(\lambda)$ random variables are independent exponential random variables with parameters depending on the dimension n and λ. Specifically, for i.i.d. $X_i \sim \text{Exp}(\lambda)$ the spacings become independent W_i's that are distributed as

$$W_i \sim \text{Exp}(\lambda(n-i)),\ \forall i \in [1:n-1]. \tag{C.27}$$

It then follows that

$$f_{W_i}(0^+) = \lim_{w\to 0^+} \lambda(n-i)\mathrm{e}^{-\lambda(n-i)w} = \lambda(n-i),\ \forall i, \tag{C.28}$$

$$f_{W_i,W_j}(\mathbf{0}_2^+) = f_{W_i}(0^+) f_{W_j}(0^+) = \lambda^2(n-i)(n-j),\ \forall i \neq j. \tag{C.29}$$

It is also easy to evaluate

$$f'_{W_i}(0^+) = \lim_{w\to 0^+} -\lambda^2(n-i)^2 e^{-\lambda(n-i)w} = -\lambda^2(n-i)^2,\ \forall i. \tag{C.30}$$

By substituting (C.28) and (C.30) into c_1 and c_2 in Corollay (5.3.7) with some algebras, we conclude the proof of Example 5.3.9.

C.6 Proof of Corollary 5.3.11

Since $P_e(\phi_{\text{lin}}, \mathcal{K}) \to 0$ as $\sigma \to 0^+$, the first order rate is given from Theorem 5.3.6 by

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = P_e^{(1)}\sigma + O(\sigma^2), \tag{C.31}$$

where

$$
\begin{aligned}
P_e^{(1)} &= \alpha_1(0^+) \\
&= \lim_{\sigma\to 0^+} \sum_{\substack{\mathcal{I}\subseteq[1:n-1]\\|\mathcal{I}|=1}} \int_{\mathbf{u}\in\mathbb{R}_+^k} F_{\mathbf{V}_\mathcal{I}}(-\mathbf{u}) f_{\mathbf{W}_\mathcal{I}}(\sigma\mathbf{u})\,\mathrm{d}\mathbf{u} \\
&= \lim_{\sigma\to 0^+} \sum_{i=1}^{n-1} \int_0^\infty \Pr(V_i \le -u) f_{W_i}(\sigma u)\,\mathrm{d}u \\
&\overset{(a)}{=} \lim_{\sigma\to 0^+} \sum_{i=1}^{n-1} \int_0^\infty \Pr(V_1 \ge u) f_{W_i}(\sigma u)\,\mathrm{d}u \\
&\overset{(b)}{=} \sum_{i=1}^{n-1} f_{W_i}(0^+) \int_0^\infty \Pr(V_1 \ge u)\,\mathrm{d}u \\
&= \sum_{i=1}^{n-1} f_{W_i}\left(0^+\right) \int_0^\infty \mathbb{E}\left[\mathbb{1}_{\{V_1>u\}}\right]\,\mathrm{d}u \\
&\overset{(c)}{=} \sum_{i=1}^{n-1} f_{W_i}\left(0^+\right) \mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{V_1>u\}}\,\mathrm{d}u\right] \\
&= \sum_{i=1}^{n-1} f_{W_i}\left(0^+\right) \mathbb{E}\left[\int_0^{\max\{V_1,0\}} 1\,\mathrm{d}u\right] \\
&= \sum_{i=1}^{n-1} f_{W_i}\left(0^+\right) \mathbb{E}\left[\max\{V_1,0\}\right] \\
&\overset{(d)}{=} \frac{1}{2}\sum_{i=1}^{n-1} f_{W_i}(0^+)\mathbb{E}\left[|V_1|\right], \qquad \text{(C.32)}
\end{aligned}
$$

where the labeled equalities follow from: (a) the exchangeability of $\mathbf{N}$ (i.e., $V_i = N_{i+1} - N_i \overset{d}{=} N_i - N_{i+1} = -V_i$ and $V_i \overset{d}{=} V_j,\ \forall(i,j)$; (b) the dominated convergence theorem; (c) the Fubini-Tonelli theorem; and (d) the symmetry of V as we have used in step (a). We conclude the proof of Corollary 5.3.11 by substituting (C.32) into (C.31).

C.7 Proof of Proposition 5.3.14

For i.i.d. $X_i \sim F_X$, where F_X is the CDF of X, we observe that [60]

$$\begin{aligned}
\sum_{i=1}^{n-1} f_{W_i}(0^+) &= \sum_{i=1}^{n-1} \frac{n!}{(i-1)!(n-i-1)!} \int_{-\infty}^{\infty} (F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&\overset{(a)}{=} \int_{-\infty}^{\infty} \sum_{i=1}^{n-1} \frac{n!}{(i-1)!(n-i-1)!} (F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sum_{i=1}^{n-1} \binom{n}{i-1} (n-i+1)(n-i)(F_X(x))^{i-1}(1-F_X(x))^{n-i-1} f_X^2(x)\,\mathrm{d}x \\
&\overset{(b)}{=} \int_{-\infty}^{\infty} \sum_{j=0}^{n-2} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^{j}(1-F_X(x))^{n-j-2} f_X^2(x)\,\mathrm{d}x,
\end{aligned} \tag{C.33}$$

where (a) follows by using the Fubini-Tonelli theorem, and (b) follows from the change of variable $j = i - 1$. To simplify the integrand in (C.33), we make use of the following,

$$\begin{aligned}
&\sum_{j=0}^{n-2} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^{j}(1-F_X(x))^{n-j-2} \\
&= \sum_{j=0}^{n} \binom{n}{j} (n-j)(n-j-1)(F_X(x))^{j}(1-F_X(x))^{n-j-2} \\
&\overset{(c)}{=} \mathbb{E}\left[(n-B)(n-B-1)\right](1-F_X(x))^{-2} \\
&= \mathbb{E}\left[n^2 - 2nB + B^2 - n + B\right](1-F_X(x))^{-2} \\
&= (n^2 - 2n^2F_X(x) + nF_X(x)(1-F_X(x)) + n^2F_X^2(x) - n + nF_X(x))(1-F_X(x))^{-2} \\
&= n(n - 2nF_X(x) + 2F_X(x) - F_X^2(x) + nF_X^2(x) - 1)(1-F_X(x))^{-2} \\
&= n(n-1)(1 - 2F_X(x) + F_X^2(x))(1-F_X(x))^{-2} \\
&= n(n-1),
\end{aligned} \tag{C.34}$$

where in (c) we let $B \sim \mathrm{Bin}(n, F_X(x))$ be the binomial random variable with parameters n and $F_X(x)$, and the expectation is with respect to B.

Then, we have

$$\sum_{i=1}^{n-1} f_{W_i}(0^+) = \int_{-\infty}^{\infty} n(n-1) f_X^2(x) \mathrm{d}x$$
$$= n(n-1) \int_{-\infty}^{\infty} f_X^2(x) \mathrm{d}x.$$

This concludes the proof of Proposition 5.3.14.

C.8 Proof of Proposition 5.4.4

We consider the Laplace mechanism such that

$$\mathcal{K}_L(\sigma, \mathbf{X}) = \mathbf{X} + \sigma \mathbf{N},$$

where $\mathbf{N}$ is i.i.d. according to $\mathrm{Lap}(0, b)$. This result has already been shown by [17] and we present it here for completeness. By using the definition of $\mathsf{RDP}_\alpha(\mathcal{K}_L)$ in Definition 5.2.2, we obtain

$$\begin{aligned}
\mathsf{RDP}_\alpha(\mathcal{K}_L) &= \sup_{(\mathbf{X},\tilde{\mathbf{X}}) \in \mathcal{X}^2 : d_H(\mathbf{X},\tilde{\mathbf{X}}) \leq 1} D_\alpha(\mathcal{K}(\mathbf{X}) \| \mathcal{K}(\tilde{\mathbf{X}})) \\
&\overset{(a)}{=} \sup_{|x_1 - x_2| \leq \ell} D_\alpha(\mathrm{Lap}(x_1, \sigma b) \| \mathrm{Lap}(x_2, \sigma b)) \\
&= \sup_{r \in [0,\ell]} D_\alpha(\mathrm{Lap}(0, \sigma b) \| \mathrm{Lap}(r, \sigma b)) \\
&\overset{(b)}{=} \sup_{r \in [0,\ell]} \frac{1}{\alpha - 1} \ln \frac{\alpha e^{-(1-\alpha) r/(\sigma b)} - (1-\alpha) e^{-\alpha r/(\sigma b)}}{2\alpha - 1} \\
&\overset{(c)}{=} \frac{1}{\alpha - 1} \ln \frac{\alpha e^{-(1-\alpha) \ell/(\sigma b)} - (1-\alpha) e^{-\alpha \ell/(\sigma b)}}{2\alpha - 1},
\end{aligned} \tag{C.35}$$

where the labeled equalities follow from: (a) the fact that $\mathcal{K}(\mathbf{X})$ and $\mathcal{K}(\tilde{\mathbf{X}})$ have nearly identical distributions that differ at only one coordinate; (b) the closed-form expression by [79]; and (c) the fact that $D_\alpha(\mathrm{Lap}(0, \sigma b) \| \mathrm{Lap}(r, \sigma b))$ is an increasing function in r. Therefore, $\mathcal{K}_L(\sigma, \mathbf{X})$ gives (α, ϵ)-RDP with ϵ given in (C.35).

For the upper and lower bounds, we first obtain an upper bound by using the concavity

property of the logarithm and its first-order condition [87], i.e.,

$$\ln(x+y) \leq \ln(x) + \frac{y}{x},\ \forall y, x > 0. \tag{C.36}$$

Using the above inequality, we can upper bound (C.35) as

$$\begin{aligned}
\epsilon &= \frac{1}{\alpha-1} \ln \frac{\alpha e^{-(1-\alpha)\ell/(\sigma b)} - (1-\alpha)e^{-\alpha\ell/(\sigma b)}}{2\alpha-1} \\
&\leq \frac{1}{\alpha-1} \left(\ln \frac{\alpha e^{(\alpha-1)\ell/(\sigma b)}}{2\alpha-1} + \frac{(\alpha-1)e^{-\alpha\ell/(\sigma b)}}{\alpha e^{(\alpha-1)\ell/(\sigma b)}} \right) \\
&= \frac{1}{\alpha-1} \ln \frac{\alpha e^{(\alpha-1)\ell/(\sigma b)}}{2\alpha-1} + \frac{1}{\alpha} e^{-\frac{(2\alpha-1)\ell}{\sigma b}} \\
&= \frac{\ell}{\sigma b} + \frac{1}{\alpha-1} \ln \frac{\alpha}{2\alpha-1} + \frac{1}{\alpha} e^{-\frac{(2\alpha-1)\ell}{\sigma b}}.
\end{aligned} \tag{C.37}$$

A lower bound can be obtained by dropping the second exponential term in (C.35) as

$$\begin{aligned}
\epsilon &\geq \frac{1}{\alpha-1} \ln \frac{\alpha e^{-(1-\alpha)\ell/(\sigma b)}}{2\alpha-1} \\
&= \frac{\ell}{\sigma b} + \frac{1}{\alpha-1} \ln \frac{\alpha}{2\alpha-1}.
\end{aligned} \tag{C.38}$$

This concludes the proof of Proposition 5.4.4.

C.9 Proof of Corollary 5.4.5

From the lower bound in (5.28) in Proposition 5.4.4, we directly obtain the lower bound of σ as

$$\frac{\ell}{b\left(\epsilon + \frac{1}{\alpha-1} \ln \frac{2\alpha-1}{\alpha}\right)} \leq \sigma. \tag{C.39}$$

In addition, the upper bound in (5.28) can be further bounded by

$$\epsilon \leq \frac{\ell}{\sigma b}, \tag{C.40}$$

which follows from the fact that $\frac{1}{\alpha-1} \ln \frac{\alpha}{2\alpha-1} + \frac{1}{\alpha} e^{-\frac{(2\alpha-1)\ell}{\sigma b}}$ is increasing in $\alpha > 1$ for any values

of $\sigma > 0, b > 0$, and $\ell \geq 0$, and the limit is 0, i.e.,

$$\lim_{\alpha\to\infty} \frac{1}{\alpha-1}\ln\frac{\alpha}{2\alpha-1} + \frac{1}{\alpha}e^{-\frac{(2\alpha-1)\ell}{\sigma b}} = 0.$$

The bound in (C.40) gives the upper bound on σ as

$$\sigma \leq \frac{\ell}{b\epsilon}, \tag{C.41}$$

and hence, with (C.39) we have that, for the Laplace mechanism,

$$\frac{\ell}{b\left(\epsilon + \frac{1}{\alpha-1}\ln\frac{2\alpha-1}{\alpha}\right)} \leq \mathsf{RDP}_\alpha^{-1}(\epsilon,\ell) \leq \frac{\ell}{b\epsilon}. \tag{C.42}$$

We now leverage Proposition 5.4.2, which requires $\mathbb{E}[|V|]$. In order to have $\mathrm{Var}(N) = 1$ for $N \sim \mathrm{Lap}(0,b)$, we set $b = \frac{1}{\sqrt{2}}$, and we evaluate $\mathbb{E}[|V|]$ as follows. The PDF of $V = N_1 - N_2$, where N_1 and N_2 are independent Lap $\left(0, \frac{1}{\sqrt{2}}\right)$, is given by

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{N_1}(z) f_{N_2}(v-z)\,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2}}e^{-\sqrt{2}|z|}\frac{1}{\sqrt{2}}e^{-\sqrt{2}|v-z|}\,\mathrm{d}z \\ &= \frac{1}{2\sqrt{2}}e^{-\sqrt{2}|v|} + \frac{1}{2}|v|e^{-\sqrt{2}|v|}. \end{aligned} \tag{C.43}$$

Then, by the symmetry of V we have

$$\mathbb{E}[|V|] = 2\int_0^\infty v f_V(v)\,\mathrm{d}v = \frac{3}{2\sqrt{2}}. \tag{C.44}$$

We obtain the trade-off expression by combining (C.42) and (5.25) as

$$P_e(\phi_{\text{lin}}, \mathcal{K}) = \frac{3C_{\mathbf{X}}}{4\sqrt{2}}\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell) + O\left(\frac{1}{\epsilon^2}\right), \tag{C.45}$$

where

$$\frac{\sqrt{2}\ell}{\epsilon + \frac{1}{\alpha-1}\ln\frac{2\alpha-1}{\alpha}} \leq \mathsf{RDP}_\alpha^{-1}(\epsilon,\ell) \leq \frac{\sqrt{2}\ell}{\epsilon}. \tag{C.46}$$

This concludes the proof of Corollary 5.4.5.

C.10 Proof of Proposition 5.4.8

Consider the Gaussian mechanism such that

$$\mathcal{K}_G(\sigma, \mathbf{X}) = \mathbf{X} + \sigma\mathbf{N},$$

where $\mathbf{N}$ is i.i.d. according to $\mathcal{N}(0,1)$. By using the definition of $\mathsf{RDP}_\alpha(\mathcal{K}_G)$ in Definition 5.2.2, we obtain

$$\begin{aligned}
\mathsf{RDP}_\alpha(\mathcal{K}_G) &= \sup_{(\mathbf{X},\tilde{\mathbf{X}})\in\mathcal{X}^2 : d_H(\mathbf{X},\tilde{\mathbf{X}})\leq 1} D_\alpha(\mathcal{K}(\mathbf{X})\|\mathcal{K}(\tilde{\mathbf{X}})) \\
&\overset{(a)}{=} \sup_{|x_1-x_2|\leq \ell} D_\alpha(\mathcal{N}(x_1,\sigma^2)\|\mathcal{N}(x_2,\sigma^2)) \\
&= \sup_{r\in[0,\ell]} D_\alpha(\mathcal{N}(0,\sigma^2)\|\mathcal{N}(r,\sigma^2)) \\
&\overset{(b)}{=} \sup_{r\in[0,\ell]} \frac{1}{2}\frac{\alpha r^2}{\sigma^2} \\
&\overset{(c)}{=} \frac{1}{2}\frac{\alpha\ell^2}{\sigma^2},
\end{aligned} \tag{C.47}$$

where the labeled equalities follow from: (a) the fact that $\mathcal{K}(\mathbf{X})$ and $\mathcal{K}(\tilde{\mathbf{X}})$ have nearly identical distributions that differ at only one coordinate; (b) the closed-form expression by [79]; and (c) the fact that $\frac{\alpha r^2}{\sigma^2}$ is an increasing function in r. Using (5.24), we obtain

$$\mathsf{RDP}_\alpha^{-1}(\epsilon,\ell) = \sqrt{\frac{\alpha\ell^2}{2\epsilon}}. \tag{C.48}$$

This concludes the proof of Proposition 5.4.8.

C.11 Proof of Proposition 5.4.11

We start by noting that σN with $N \sim \mathcal{GN}(0, h(p), p)$ has variance equal to σ^2 and PDF given by [75],

$$f_{\sigma N}(z) = K \exp\left(-\left|\frac{z}{\sigma h(p)}\right|^p\right), \tag{C.49}$$

where $h(p) = \sqrt{\frac{\Gamma(p^{-1})}{\Gamma(3p^{-1})}}$ and $K = \frac{p}{2\sigma h(p)\Gamma(p^{-1})}$.

Since (∞, ϵ)-RDP is equivalent to ϵ-DP, we evaluate the Rènyi divergence of order $\alpha = \infty$. From Definition 5.2.2, the Rényi divergence of order $\alpha = \infty$ between $\mathbf{x} + \sigma\mathbf{N}$ and $\tilde{\mathbf{x}} + \sigma\mathbf{N}$ with $d_H(\mathbf{x}, \tilde{\mathbf{x}}) \leq 1$ is

$$\begin{aligned} D_\infty(\mathcal{K}(\mathbf{x} + \sigma\mathbf{N}) \| \mathcal{K}(\tilde{\mathbf{x}} + \sigma\mathbf{N})) &\overset{(a)}{=} D_\infty(\mathcal{K}(r + \sigma N) \| \mathcal{K}(\sigma N)) \\ &\overset{(b)}{=} \sup_{z \in \mathbb{R}} \log \frac{f_{\sigma N}(z - r)}{f_{\sigma N}(z)} \\ &= \sup_{z \in \mathbb{R}} \left\{ -\left|\frac{z - r}{\sigma h(p)}\right|^p + \left|\frac{z}{\sigma h(p)}\right|^p \right\} \\ &\overset{(c)}{=} \left|\frac{r}{\sigma h(p)}\right|^p, \end{aligned} \tag{C.50}$$

where the labeled equalities follow from: (a) the fact that $r \in [-\ell, \ell]$ is the difference between $\mathbf{x}$ and $\tilde{\mathbf{x}}$, and without loss of generality we consider positive $r \in [0, \ell]$ due to the symmetry of N; (b) the definition of the Rényi divergence of order $\alpha = \infty$; and (c) the fact that the maximum of the function $t \mapsto |t|^p - |t - r|^p$ is obtained at $t = r$ for $0 < p < 1$.

Since (C.50) is an increasing function in $r \in [0, \ell]$, we obtain

$$\mathsf{RDP}_\infty(\mathcal{K}_{GN}) = \sup_{r \in [0, \ell]} \left|\frac{r}{\sigma h(p)}\right|^p = \frac{1}{\sigma^p}\left(\frac{\ell}{h(p)}\right)^p. \tag{C.51}$$

This concludes the proof of Proposition 5.4.11.

C.12 Minimizing (5.35) with respect to $0 < p \leq 1$

To find the minimum value (or minimizer) of $\frac{1}{h(p)}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}$ with respect to $0 < p \leq 1$, we differentiate it with respect to p and obtain

$$\begin{aligned}
\frac{\partial}{\partial p}\left\{\frac{1}{h(p)}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}\right\} &= \frac{\partial}{\partial p}\left\{\sqrt{\frac{\Gamma(3p^{-1})}{\Gamma(p^{-1})}}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}\right\} \\
&= \frac{\partial p^{-1}}{\partial p}\frac{\partial}{\partial p^{-1}}\left\{\sqrt{\frac{\Gamma(3p^{-1})}{\Gamma(p^{-1})}}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}\right\} \\
&\overset{\text{(a)}}{=} \frac{\partial p^{-1}}{\partial p}\frac{\partial}{\partial x}\left\{\sqrt{e^{\ln\Gamma(3x)-\ln\Gamma(x)}}\left(\frac{1}{\epsilon}\right)^{x}\right\} \\
&\overset{\text{(b)}}{=} -\frac{1}{p^2}\frac{(3\psi(3x)-\psi(x)-2\ln\epsilon)}{2}\sqrt{\frac{\Gamma(3x)}{\Gamma(x)}}\left(\frac{1}{\epsilon}\right)^{x} \\
&= -\frac{1}{2p^2}\sqrt{\frac{\Gamma(\frac{3}{p})}{\Gamma(\frac{1}{p})}}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}\left(3\psi(3p^{-1})-\psi(p^{-1})-2\ln\epsilon\right), \qquad \text{(C.52)}
\end{aligned}$$

where the labeled equalities follow from: (a) the change of variable $x = p^{-1}$; and (b) letting $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ be the digamma function [88, p.258].

By using the change of variable $x = p^{-1} \in [1, \infty)$, we have an equivalent expression for the derivative,

$$-\frac{x^2}{2}\sqrt{\frac{\Gamma(3x)}{\Gamma(x)}}\left(\frac{1}{\epsilon}\right)^{x}(3\psi(3x)-\psi(x)-2\ln\epsilon). \qquad \text{(C.53)}$$

Since the sign of $-\frac{x^2}{2}\sqrt{\frac{\Gamma(3x)}{\Gamma(x)}}\left(\frac{1}{\epsilon}\right)^{x}$ is negative for all $x \in [1, \infty)$, it is sufficient to consider the last term $(3\psi(3x) - \psi(x) - 2\ln\epsilon)$. The digamma function $\psi(x)$ does not have a closed-form expression and hence, we instead use the approximation $\psi(x) \approx \ln x - \frac{1}{cx}$, where $1 \leq c \leq 2$ is a constant. This approximation expression comes from the following bounds [89, Lemma 2] for $x \geq 1$,

$$\ln x - \frac{1}{x} \leq \psi(x) \leq \ln x - \frac{1}{2x}. \qquad \text{(C.54)}$$

Using $\psi(x) \approx \ln x - \frac{1}{cx}$, we have that

$$3\psi(3x) - \psi(x) - 2\ln\epsilon \approx \ln(27x^2) - \ln\epsilon^2, \tag{C.55}$$

which is increasing in x. Hence, (C.53) is negative if $\frac{\epsilon}{3\sqrt{3}} < x$ and is positive otherwise, which implies that the minimum value of $\frac{1}{h(p)}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}$ can be obtained by choosing p such that $\frac{1}{p} = x \approx \frac{\epsilon}{3\sqrt{3}}$. Due to the condition $p \leq 1$, we finally obtain the approximed minimizer p for $\frac{1}{h(p)}\left(\frac{1}{\epsilon}\right)^{\frac{1}{p}}$ given by

$$\hat{p} = \min\left\{\frac{3\sqrt{3}}{\epsilon}, 1\right\}. \tag{C.56}$$

Note that an exact expression for the minimizer is $p = \min\{p', 1\}$ where p' is such that

$$3\psi\left(\frac{3}{p'}\right) - \psi\left(\frac{1}{p'}\right) - 2\ln\epsilon = 0, \tag{C.57}$$

which can be obtained numerically.