DETECTION WITH PARTIAL INFORMATION FOR THE GAUSSIAN SETUP IN THE POTENTIAL PRESENCE OF A JAMMER

by

Onur Özyeşil

B. S., in Electrical and Electronics Engineering, Boğaziçi University, 2006

B. S., in Mathematics, Boğaziçi University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Master of Science in Electrical and Electronics Engineering

Boğaziçi University

2008

ACKNOWLEDGEMENTS

This thesis could not have been written without the guidance, patience and intelligence of my advisor Prof. Kıvanç Mıhçak who not only made it possible for me to study a meaningful and theoretically satisfactory problem during my graduate studies but also helped me to comprehend the structure of an academic study and get the pleasure of constructing and solving a valuable problem. I consider myself lucky for being a graduate student of Boğaziçi University in a period when Prof. Mıhçak is with our department. Additionally, I want to thank all my friends in graduate school for their spiritual and academic supports and for not leaving me alone in the long nights of mine spent for studying.

On the other hand, at least half of the effort spent for this thesis is due to my family in Istanbul and my family in Manisa. Among them all, I should confess that, my desire for academic studies and consequently this thesis might not even exist without the huge love I have for my Özlem, who is the untouched piece of my soul from where, I believe, all the source of my will to live arises.

Also, I want to thank TÜBITAK for financially supporting me via the full Graduate Studies Scholarship, No. 2228.

ABSTRACT

DETECTION WITH PARTIAL INFORMATION FOR THE GAUSSIAN SETUP IN THE POTENTIAL PRESENCE OF A JAMMER

We introduce the problem of communication with partial information, where there is an asymmetry between the transmitter and the receiver codebooks. We study this setup in a binary detection theoretic context for the additive colored Gaussian noise channel in the potential presence of a jammer. In our proposed setup, the partial information available at the detector consists of dimensionality-reduced versions of the transmitter codewords, where the dimensionality reduction is achieved via a linear transform. In the first part of the thesis, we focus on the "no-jammer" case and accordingly find the MAP-optimal detection rule and the corresponding conditional probability of error (conditioned on the partial information the detector possesses). Then, we constructively quantify two optimal classes of linear transforms: For the first class, the cost function is the expected Chernoff bound on the conditional probability of error of the MAP-optimal detector; for the second class, the cost function is a certain upper bound on the failure probability, which is defined as the probability of the aforementioned conditional error probability being greater than a given constant. In the second part of the thesis, we study the case where an active jammer is present (subject to a peak power constraint) together with additive colored Gaussian noise. In this case, we first derive the conditional probability of error of a minimum Euclidean distance detector as a function of the receiver partial information and the jammer signal. Then, we quantify the worst-case jammer strategy, which maximizes the aforementioned conditional probability of error. As a result, we propose a criterion for choosing the dimensionality-reducing linear transforms in the sense of worst-case failure probability.

ÖZET

POTANSİYEL BİR KARIŞTIRICININ VARLIĞINDA GAUSS KURGUSU İÇİN KISMİ BİLGİYLE SEZİM

Tezimizde verici ve alıcı kod çizelgeleri arasında bir bakışımsızlığın bulunduğu "kısmi bilgiyle iletişim" problemini sunuyoruz. Bu yapıyı potansiyel bir karıştırıcının varlığında Gauss dağılımına sahip gürültü kanalı için sezim-teoriksel bir içerikte inceliyoruz. Öngörulen yapıda, alıcı tarafında mevcut olan kısmi bilgi, verici kod çizelgesinin boyutlarının azaltılarak doğrusal bir dönüştürücü aracılığıyla gerçekleştirilen bir uyarlamasından oluşmaktadır. Tezin ilk kısmında, karıştırıcının var olmadığı duruma odaklanarak maksimum-sonsal-olasılık açısından en iyilenmiş sezim kuralı ve bu kurala dayanan koşullu (kısmi bilgiye koşullanmış) hata olasılığını türetiyoruz. Sonrasında, yapıcı bir şekilde iki tür en iyilenmiş doğrusal dönüşüm sınıfını niceliyoruz: Birinci sınıfta zarar fonksiyonu, maksimum-sonsal-olasılık açısından en iyilenmiş sezicinin koşullu hata olasılığı üzerindeki Chernoff sınırının beklenen değerinden; ikinci sınıfta ise, önceden belirttiğimiz koşullu hata olasılığının verili bir değerden daha büyük olma olasılığı olarak tanımladığımız, "başarısızlık olasılığı" üzerinde bilinen bir sınırdan oluşmaktadır. Tezin sonraki kısmında, karıştırıcının eklenen Gauss gürültüsüyle birlikte mevcut bulunduğu (en fazla güç sınırlamasıyla) durumu incelemekteyiz. Bu durumda, öncelikle en-küçük-Öklit-uzaklığı sezicisinin koşullu hata olasılığını, kısmi bilginin ve karıştırıcının işaretinin bir fonksiyonu olarak türetiyoruz. Daha sonra, belirtilen koşullu hata olasılığını en çok arttıran, "en-kötü durum" karıştırıcı yaklaşımını niceliyoruz. Sonuç olarak, en kötü durum başarısızlık olasılığı açısından doğrusal dönüştürücülerin seçimi için bir ölçüt öneriyoruz.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

$\mathbf{e}$	Additive Gaussian channel noise
i	Binary information
$J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$	Jammer function
m	Partial information size
n	Transmitted signal size
P_e	Expected probability of error
P_α	Failure probability
P_{α,J^*}	Worst-case failure probability
$\mathbf{T}$	Linear transformation for partial information extraction
$\mathbf{u}$	Jammer signal
$\mathbf{x}_i$	Transmitter codeword corresponding to binary information i
$\mathbf{y}$	Channel output
$\mathbf{z}_i$	Detector codeword corresponding to binary information i
α	Conditional probability of error threshold
$\mathcal{N}(\mu, \mathbf{\Sigma})$	Multivariate Gaussian distribution with mean μ and covariance $\mathbf{\Sigma}$
$\mathcal{S}_\mathbf{M}$	Set of orthonormal $n \times m$ real-valued matrices
$\mathcal{S}_\mathbf{T}$	Set of $m \times n$ real-valued matrices of rank m
$\mathbf{\Sigma}_e$	Covariance matrix of noise
$\mathbf{\Sigma}_x$	Covariance matrix of transmitted signals
$\mathbf{\Sigma}_z$	Covariance matrix of partial information

a.s.	Almost surely
i.i.d.	Independent and identically distributed
SVD	Singular value decomposition

1. INTRODUCTION

In this thesis, we introduce a communication-theoretic paradigm, which we name as "communication with partial information", and subsequently study it within a detection-theoretic context (therefore the term "detection with partial information") in the particular case of Gaussian setup, where the presence of a potential jammer is also incorporated. In the proposed paradigm, there is an inherent asymmetry between the information the transmitter and the receiver possess in terms of the utilized codebooks. In particular, in the "detection with partial information" setup, the codebook of the receiver is formed via applying a non-invertible process on the codebook of the transmitter; hence *the codebooks are different*. Thus, the information available at the transmitter forms a "superset" of the information available at the receiver. Note that, a reminiscent asymmetric structure between the transmitter and the receiver also exists in the well-known family of problems, termed as "communication with side information" [1, 2, 3, 4]. However, in the paradigm of "communication with side information" (unlike the proposed "communication with partial information" setup), the utilized codebooks at the receiver and the transmitter are the same; in addition, either the transmitter or the receiver is "favored" with the presence of "extra" information (which amounts to the "side information").

It appears that, there are at least two significant applications that motivate the formulation of the "communication with partial information" approach:

- The first application can be viewed to fall within the category of "robust signal hashing" in the signal processing & multimedia security literature [5, 6, 7, 8]. In robust signal hashing, a content owner provides "robust hash value"s of the protected content (that is some dimensionality-reduced versions of the protected content) to a third party, which searches the content using its robust hash values as *the partial information* at the receiver end. These robust hash values represent "the content's significant features" and are ideally approximately-invariant under acceptable modifications to the content . In practical applications, the third

party that performs the hash-based search is usually *not trusted*; hence, there is a significant issue of privacy. In particular, given a robust hash value, it should ideally be impossible to retrieve the original protected content from a privacy viewpoint. The setup proposed in this thesis can be used as a detection-theoretic model to analyze the hash-based detection problem: the protected content is represented by the transmitted signal; the robust hash values used in the search are represented by the partial information available at the receiver; a perceptually-acceptable modification to the protected content is represented by the channel noise.

- The second application includes all instances of point-to-point communications, where there is an inherent asymmetry between the transmitter and the receiver in terms of their storage capabilities and computational resources. In particular, the cases, when the receiver is unable to store the codebook used by the encoder (due to a limit on the memory) or utilize the codebook used by the encoder (due to a limit on the computational resources), can be studied within the framework of "communication with partial information". In such cases, one potential remedy is the receiver's using a "simplified" (i.e., dimensionality-reduced) version of the codebook of the encoder. In practice, such situations may typically arise, for instance, when there is a bi-directional communication between a sensor and the base station (the resource-limited receiver representing the sensor) or when there is a bi-directional communication between a controller and a remote measurement unit. In such applications, the simplified version of the encoder codebook is represented by the partial information at the receiver side.

The aforementioned application scenarios may just physically appear in various contexts, in which "the worst case performance" is a valid concern. In order to address this issue, in the most general sense, we formulate the problem such that the presence of an "all-knowing" intelligent jammer, who acts so as to degrade the receiver performance as much as possible subject to some power constraint, is incorporated. For studies on jammers in related (yet different) communication theoretic problems, see, for example, [9, 10, 11, 12] and the references therein. In our formulation, the jammer possesses all the required information, in the sense that it has access to both transmitter and

receiver codebooks and also the transmitted message bit. Then, the jammer introduces an extra additive distortion to the transmitter output subject to a peak power constraint, subsequently followed by an additive (colored) Gaussian channel, which occurs due to natural circumstances. As such, this setup yields the worst-case performance of the proposed detection with partial information setup. Furthermore, the presence of a jammer also addresses the practical scenario of "robust signal hashing" where security is a significant concern; here the jammer represents an "all-knowing" adversary. Note that, the special case of imposing a trivially-zero peak power constraint on the jammer reduces to the additive (colored) Gaussian channel case discussed above. Our contributions in this thesis can be listed as follows:

- We introduce the paradigm of "communication with partial information" and study it within the context of binary detection in the Gaussian setup. We believe the main philosophy behind this formulation (i.e., introducing an asymmetry between the transmitter and the receiver in the sense of utilized codebooks) can be used to analyze various problems of interest in communication theory and signal processing.
- Within the binary hypothesis testing setup, we study the "no-jammer" case, where the disturbance on the transmitter output consists only of additive colored Gaussian noise, and the detector partial information is produced via applying a linear (dimensionality-reducing) transform on the encoder codebook. We derive several results of interest given below:
  - We derive the MAP-optimal detection rule and the corresponding probability of error, both of which are conditioned on the partial information available at the detector.
  - We construct a class of *optimal* linear transforms, which minimize the expected (with respect to the joint distribution of the detector partial information) Chernoff bound on the aforementioned probability of detection error.
  - We quantify the "failure probability" of the MAP-optimal detector (which is the probability of the aforementioned conditional error probability being greater than a given constant), and subsequently construct a class of *optimal* linear transforms, which minimize a certain bound on the failure probability.

- Within the binary hypothesis testing setup, we study the "active jammer" case, where the disturbance on the transmitter output consists of both additive jammer signal and additive colored Gaussian noise, and the detector partial information is produced via applying a linear (dimensionality-reducing) transform on the encoder codebook. In this case, we assume that the detector does not have full-knowledge about the jammer function, and hence applies a suboptimal, however less parametric, detection method based on "minimum Euclidean distance" in the range space of the aforementioned linear transform. The results of our study are provided below:
  - We derive the conditional probability of error of the aforementioned detector as a function of the jammer signal.
  - We characterize the class of optimal jammer functions, which maximize the aforementioned conditional probability of error at the detector.
  - In the light of the aforementioned results, we quantify a criterion in the sense of "worst-case failure probability", which is the probability of the conditional error probability being greater than a given constant in case of worst-case jammer behavior.

In Sec. 2, we provide the notation that is used throughout the thesis and specify the formal problem statement. In Sec. 3 and Sec. 4, we present our results (mentioned above) for the additive Gaussian channel case (i.e., no jammer) and the active jammer case, respectively. We conclude with final discussions of the results in Sec. 5.

2. NOTATION AND PROBLEM STATEMENT

In this section we provide the notation we use and give the general statement of the problem.

2.1. Notation

Boldface lowercase and uppercase letters denote vectors and matrices, respectively; the corresponding regular letters with subscripts denote their individual elements. For instance, given a vector $\mathbf{a}$, a_i represents its i-th element; given a matrix $\mathbf{A}$, A_{ij} denotes its (i,j)-th element. Note that, we do not use a separate notation for random vectors; we assume that it is clear from the context.

Given a matrix $\mathbf{A}$, $\mathbf{A}^T$, $r(\mathbf{A})$ and $\det(\mathbf{A})$ denote its transpose, rank and determinant, respectively; further, $\mathbf{I}_n$ denotes the identity matrix of size $n \times n$. Given the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, $\langle \mathbf{x}, \mathbf{y} \rangle$ indicates the inner product that induces the Euclidean norm, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i$; accordingly the induced Euclidean norm is denoted by $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$.

We provide a definition that will be used throughout the thesis.

Definition 2.1.1 *Given* $\mathbf{A} \in \mathbb{R}^{m \times n}$*, such that* $r(\mathbf{A}) = k$*,* Singular Value Decomposition *(SVD) of* $\mathbf{A}$ *is unique (up to ordering) and defined as*

$$\mathbf{A} \stackrel{\triangle}{=} \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \tag{2.1}$$

where $\mathbf{U} \in \mathbb{R}^{m \times k}$*,* $\mathbf{V} \in \mathbb{R}^{n \times k}$*,* $\mathbf{\Lambda} \in \mathbb{R}^{k \times k}$ *are called the left-singular vector matrix, the right-singular vector matrix and the singular value matrix of* $\mathbf{A}$*, respectively. The matrix* $\mathbf{\Lambda}$ *is diagonal and these diagonal elements are termed the singular values of* $\mathbf{A}$*.*

Given the matrix $\mathbf{A} \in \mathbb{R}^{m\times k}$ of rank $r \leq \min(m,k)$, $\{\sigma_i(\mathbf{A})\}_{i=1}^{r}$ denote its non-zero singular values, which are assumed to be in non-increasing order without loss of generality. For a square matrix $\mathbf{A}$ of size $k \times k$ and of rank $r \leq k$, $\{\lambda_i(\mathbf{A})\}_{i=1}^{r}$ denote its non-zero eigenvalues; in case $\mathbf{A}$ is a symmetric matrix, $\{\lambda_i\}$ are assumed to be in non-decreasing order. We use $\mathcal{N}(\mu, \mathbf{\Sigma})$ to denote a multivariate Gaussian distribution, with mean vector μ and covariance matrix $\mathbf{\Sigma}$. Furthermore, $\mathrm{Q}(\cdot)$ denotes the standard Q-function: $\mathrm{Q}(\alpha) \triangleq \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

2.2. Problem Statement

We analyze a binary communication system, where the encoder selects one of the two codewords, $\mathbf{x}_0$ and $\mathbf{x}_1$, representing the message bit $i \in \{0,1\}$, where $\Pr(i=0) = \Pr(i=1) = 1/2$; the selected codeword, $\mathbf{x} = \mathbf{x}_i$, is sent through a channel. The encoder output $\mathbf{x}$ is first corrupted by an additive *jammer*, where $\mathbf{u}$ denotes the jamming signal. This is subsequently followed by the addition of a signal-independent (not necessarily white) Gaussian noise, denoted by $\mathbf{e}$, thereby yielding the overall channel output $\mathbf{y}$. Observing $\mathbf{y}$, the receiver acts as a detector and makes a binary decision, as to the origins of received signal. We pursue a detection-theoretic approach to solve this problem and assume uniform costs. See Fig. 2.1 for a schematic illustration of the proposed problem. We assume that $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{u}$, $\mathbf{e}$, and $\mathbf{y}$ are all length-n real-

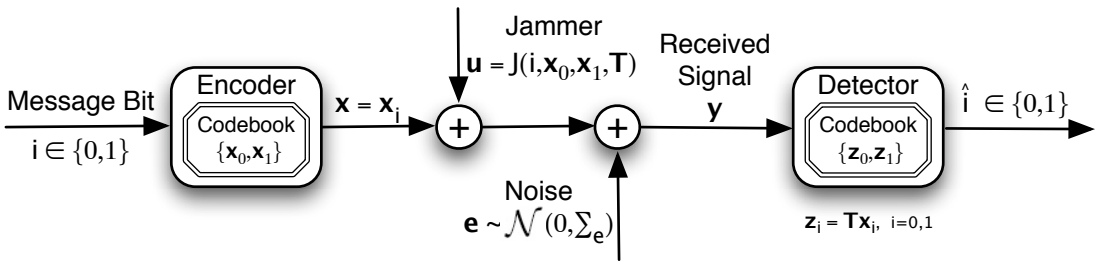

Figure 2.1. Block diagram representation of the problem of "binary detection with partial information".

valued vectors, where $\mathbf{x}_0$ and $\mathbf{x}_1$ are independent of each other and $\mathbf{x}_0, \mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_x)$, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_e)$ is independent of both $\mathbf{x}_0$ and $\mathbf{x}_1$. Here, we also assume that the covariance matrices of the original signals $\mathbf{\Sigma}_x$ and the covariance matrix of the noise $\mathbf{\Sigma}_e$

are *positive definite* (where we already know that they are also symmetric matrices). Note that, the proposed problem is radically different from the conventional binary detection scenario due to the *mismatch between the codebooks of the encoder and the detector*. In particular, *the detector does not know the original codewords* $\{\mathbf{x}_0, \mathbf{x}_1\}$*, but only their distributions and their dimensionality-reduced versions,* $\{\mathbf{z}_0, \mathbf{z}_1\}$, where $\mathbf{z}_i = \mathbf{T} \cdot \mathbf{x}_i$, $i = 0, 1$, and $\mathbf{T}$ is a deterministic real matrix of size $m \times n$, $m < n$, $r(\mathbf{T}) = m$. Note that, this implies, $\mathbf{z}_0$ and $\mathbf{z}_1$ are both length-m real-valued vectors. As a result, our problem can be viewed as "detection with partial information" for the Gaussian case in the presence of a (potential) jammer.

The "jammer" deserves a separate discussion. First of all, the jammer is assumed to know everything the encoder knows to realize the worst-case situation; hence, the jammer function's arguments are i, $\mathbf{x}_0$, $\mathbf{x}_1$ and $\mathbf{T}$. Also, for $\mathbf{u} \stackrel{\triangle}{=} J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$ to make sense as a random vector, the jammer function $J(\cdot)$ is assumed to be a Borel-measurable function of its arguments, i, $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{T}$. Furthermore, the jammer also needs to satisfy an *peak power constraint*, which is imposed by

$$\|\mathbf{u}\|^2 \leq P \quad \text{with probability 1.} \tag{2.2}$$

Note that, our formulation allows $J(\cdot)$ to be a random function of its arguments; however, we shall soon show that at optimality, jammer chooses $J(\cdot)$ to be a deterministic function.

We analyze the aforementioned system for two important cases:

- <u>No Jammer:</u> The case of no jammer (i.e., imposing $\mathbf{u} = 0$ with probability 1) is an important class (since a jammer simply does not exist in a considerable number of situations of practical interest), which is covered in Sec. 3. An important consequence of this case is that the receiver fully knows the statistical characterization of the whole system, and hence is able to apply the MAP decoding rule. In particular, in Sec. 3.1, we derive the MAP detection rule, which is given as a function of the partial information $(\mathbf{z}_0, \mathbf{z}_1)$, and the corresponding conditional

probability of error (conditioned on $\mathbf{z}_0$ and $\mathbf{z}_1$). Subsequently, in Sec. 3.2 (resp. in Sec. 3.3), we derive the optimal linear transform, $\mathbf{T}$, in the sense of the expected Chernoff bound on the conditional probability of error (resp. in the sense of an upper bound on the cumulative distribution function of the conditional probability of error) of the MAP detector, where $\mathbf{u} = 0$ with probability 1.

- With Jammer: In this case, the receiver is assumed to be aware of a potential presence of a jammer (though it is unaware of the capabilities, and so the functional behavior of the jammer). As a result, the receiver applies the standard method of minimum Euclidean distance detection within "the partial information subspace" (i.e., the range space of $\mathbf{T}$). In Sec. 4, we investigate the optimality of the linear transform $\mathbf{T}$ in the worst case scenario in the sense of the resulting detector error probability. In particular, in Sec. 4.1, we derive the optimal jammer function $J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$ in the sense of the probability of error of the minimum Euclidean distance detector subject to the peak power constraint (2.2). In Sec. 4.2, we introduce an optimality criterion for the linear transform $\mathbf{T}$ in the presence of the optimal jammer (worst case) in the sense of the cumulative distribution of the conditional probability of error of the minimum Euclidean distance detector. Hence, the method pursued in Sec. 4 can be viewed to be analogous to solving a "minimax problem" where the jammer (resp. the designer) aims to maximize (resp. minimize) the probability of error of the detector.

Remark 2.2.1 *In [13], the authors study a closely-related problem, which can be viewed as the "deterministic variant" of the setup of Sec. 3. In particular, in [13] the authors assume that the encoder codewords $\{\mathbf{x}_i\}$ are deterministic, unknown and the subsequent analysis is based on the probability of error induced by the GLRT (generalized likelihood ratio test) rule. On the other hand, in Sec. 3 we assume that the encoder codewords $\{\mathbf{x}_i\}$ are random (in particular Gaussian) and perform a MAP-based analysis.*

Remark 2.2.2 *Although the problem imposed in this thesis is the binary detection case, the analysis can be extended to apply a "union bound based approach" for the L-*

ary case with little or no difficulty[1] *. A similar approach and discussion was provided in [13] for the case of deterministic* $\{\mathbf{x}_i\}$.

Remark 2.2.3 *In the scenario investigated in Sec. 4, since the jammer is "all-knowing", it has the advantage, i.e., the jammer designs the optimal* $\mathbf{u}$ *as a function of* $\mathbf{T}$*. The designer, on the other hand, constructs* $\mathbf{T}$ *taking into account the fact that the jammer would perform the optimal attack for any chosen linear transform. Thus, mathematically we first study an "inner maximization problem" over* $\mathbf{u}$ *(cf. Sec. 4.1), followed by an "outer minimization problem" over* $\mathbf{T}$ *where* $\mathbf{u}$ *is optimal (cf. Sec. 4.2).*

[1] In the L-ary case, the message is $\log L$ bits long; the encoder and receiver codebooks are $\{\mathbf{x}_i\}_{i=0}^{L-1}$ and $\{\mathbf{z}_i\}_{i=0}^{L-1}$, respectively.

3. DETECTION IN COLORED GAUSSIAN CHANNEL

In this section, we concentrate on the problem of binary detection with partial information (cf. Fig. 2.1), where the channel consists only of addition of colored Gaussian noise, i.e., $\mathbf{u} = 0$ almost surely. We begin our developments with deriving the optimal detection rule (conditioned on the partial information available at the detector) and find a closed-form expression for the corresponding conditional probability of error. Then, we proceed with deriving optimal linear transforms in two different senses, as explained in Sec. 2.2.

3.1. Optimal Detection Conditioned On The Partial Information

At the detector side, we are given $\{\mathbf{z}_0, \mathbf{z}_1\}$, which yield partial information about the true codewords $\{\mathbf{x}_0, \mathbf{x}_1\}$. The *binary hypothesis testing* approach on the detector side utilizes the MAP detection rule [14]: It operates on the observed data $\mathbf{y}$ (generated by the process explained in Sec. 2.2), and makes a binary decision regarding the message bit. As indicated in Sec. 2.2, we assume equal priors and uniform costs on the hypotheses. Under these assumptions, the MAP detection rule simplifies to *the maximum likelihood detection rule* [14]. Then, the MAP-detection rule can be stated in the following way.

Definition 3.1.1 *Considering the binary hypothesis testing problem, given by*

$$H_0 \;:\; \mathbf{y} \sim p\left(\mathbf{y}|H_0\right), \tag{3.1}$$

$$H_1 \;:\; \mathbf{y} \sim p\left(\mathbf{y}|H_1\right) \tag{3.2}$$

where $p(.)$ denotes the probability density function of its argument, the maximum-likelihood detection rule *is defined as*

$$p\left(\mathbf{y}|H_0\right) \underset{H_1}{\overset{H_0}{\gtrless}} p\left(\mathbf{y}|H_1\right). \tag{3.3}$$

Given the partial information $\{\mathbf{z}_0, \mathbf{z}_1\}$, the full statistical characterization of the codewords of the encoder $\{\mathbf{x}_0, \mathbf{x}_1\}$ and the additive noise $\mathbf{e}$ (provided in Sec. 2.2), Def. 3.1.1 yields the following hypothesis testing for our problem

$$H_0 \;:\; \mathbf{y} = \mathbf{x}_0 + \mathbf{e} \;\; ; \qquad \text{given } \{\mathbf{z}_0, \mathbf{z}_1\}, \tag{3.4}$$

$$H_1 \;:\; \mathbf{y} = \mathbf{x}_1 + \mathbf{e} \;\; ; \qquad \text{given } \{\mathbf{z}_0, \mathbf{z}_1\}. \tag{3.5}$$

Note that, for all $i \in \{0, 1\}$, we have

$$p\left(\mathbf{y} \,|\, H_i\right) \;=\; p\left(\mathbf{y} \mid \mathbf{x}_i \text{ is sent}, \{\mathbf{z}_0, \mathbf{z}_1\}\right) = p\left(\mathbf{x}_i + \mathbf{e} \,|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right)\big|_{\mathbf{x}_i + \mathbf{e} = \mathbf{y}} \tag{3.6}$$

$$\;=\; p\left(\mathbf{x}_i + \mathbf{e} \,|\, \mathbf{z}_i\right)\big|_{\mathbf{x}_i + \mathbf{e} = \mathbf{y}} \tag{3.7}$$

where the reduction (3.7) requires a clarification. Given any $i, j \in \{0, 1\}$, $i \neq j$, we know that $\mathbf{x}_i$ is independent of $\mathbf{x}_j$ and $\mathbf{e}$. Thus, by definition, the original σ-field induced by $\mathbf{x}_i$ is independent of the original σ-field induced by $\{\mathbf{x}_j, \mathbf{e}\}$. As a result, the sub σ-field induced by any measurable function of $\mathbf{x}_i$ is also independent of the sub σ-field induced by any measurable function of $\{\mathbf{x}_j, \mathbf{e}\}$, implying (3.7) since linear functions are Borel-measurable. Consequently, the maximum-likelihood detection rule (3.3) for our case reduces to

$$p\left(\mathbf{x}_0 + \mathbf{e} | \mathbf{z}_0\right)\big|_{\mathbf{x}_0 + \mathbf{e} = \mathbf{y}} \underset{H_1}{\overset{H_0}{\gtrless}} p\left(\mathbf{x}_1 + \mathbf{e} | \mathbf{z}_1\right)\big|_{\mathbf{x}_1 + \mathbf{e} = \mathbf{y}} . \tag{3.8}$$

Next, we provide a theorem where we specify the optimal detection rule and the corresponding conditional probability of error (conditioned on $\{\mathbf{z}_0, \mathbf{z}_1\}$).

Theorem 3.1.1 *For the case of $P = 0$ (i.e., $\mathbf{u} = \mathbf{0}$ almost surely), the maximum likelihood detection rule (3.8) is given by*

$$\| \mathbf{\Sigma}_{y|z}^{-1/2} \left(\mathbf{y} - \mu_{y_0|z_0}\right) \| \underset{H_0}{\overset{H_1}{\lessgtr}} \| \mathbf{\Sigma}_{y|z}^{-1/2} \left(\mathbf{y} - \mu_{y_1|z_1}\right) \| \tag{3.9}$$

The corresponding (conditional) probability of error (conditioned on $\mathbf{z}_0$ and $\mathbf{z}_1$) is given

by

$$\Pr\left[error\right|\left.\left\{\mathbf{z}_0,\mathbf{z}_1\right\}\right] = Q\left(\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0}-\mu_{y_1|z_1}\right)\|}{2}\right) \tag{3.10}$$

where

$$\mu_{y_i|z_i} = E(\mathbf{y}_i\,|\,\mathbf{z}_i)\big|_{\mathbf{y}_i=\mathbf{x}_i+\mathbf{e}} = \boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\mathbf{z}_i, \quad \textit{for } i\in\{0,1\}, \tag{3.11}$$

$$\boldsymbol{\Sigma}_{y|z} = Cov(\mathbf{y}_i\,|\,\mathbf{z}_i)\big|_{\mathbf{y}_i=\mathbf{x}_i+\mathbf{e},\, i=0,1} = \boldsymbol{\Sigma}_x+\boldsymbol{\Sigma}_e-\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\mathbf{T}\boldsymbol{\Sigma}_x \tag{3.12}$$

and $\boldsymbol{\Sigma}_{y|z}$ *is positive definite.*

Proof:

See Appendix A. ∎

Remark 3.1.1 *Theorem 3.1.1 in particular states that if* $\mathbf{z}_0=\mathbf{z}_1$*, conditional probability of error is* $1/2$*, which is a restatement of the fact that if there is nothing to discriminate in terms of the codebook of the detector then* $i=0$ *and* $i=1$ *are equivalent for the codebook, converting the detection process to a* fair coin tossing experiment*.*

Remark 3.1.2 *The positive definiteness of* $\boldsymbol{\Sigma}_{y|z}$*, which is shown in Appendix A, yields that if* $\mathbf{z}_0\neq\mathbf{z}_1$*, which is true almost surely, then we get* $\Pr[error\,|\,\mathbf{z}_0,\mathbf{z}_1]<1/2$*. Moreover, the argument of the Q-function is always nonnegative (this will allow us to set a tight bound on the expected probability of error, and analyze this bound in Sec. 3.2).*

With the characterization of the conditional performance of the MAP detector given by Theorem 3.1.1, we can proceed to discuss the expected, i.e. unconditional, probability of error of the MAP detector and optimal transform(s) in this sense.

3.2. Optimal Linear Operators In The Expectation Sense

In this section we study the overall performance of the case $P = 0$. The sense of overall performance for this section is expected probability of error, where by expected probability of error we mean *unconditional* probability of error. As explained in Sec. (2.2), we first define our performance criterion and discuss an analytically tractable revision on it. Then the objective is to find linear partial information extraction scheme(s) that optimizes the system performance in terms of the criterion discussed.

We first define the expected probability of error for the MAP detector.

Definition 3.2.1 *"The expected probability of error of the MAP detector", denoted by P_e, for detection with partial information in additive Gaussian channel is*

$$P_e = E_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left(\Pr\left[error \mid \{\mathbf{z}_0,\mathbf{z}_1\}\right]\right), \tag{3.13}$$

where $E_{\{\mathbf{z}_0,\mathbf{z}_1\}}(.)$ denotes expectation with respect to the joint distribution of $\mathbf{z}_0$ and $\mathbf{z}_1$.

Using this definition with the result of Theorem 3.1.1, the expected probability of error for the additive gaussian channel, i.e. P_e, is given by

$$P_e = \mathrm{E}_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left(\mathrm{Q}\left(\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0}-\mu_{y_1|z_1}\right)\|}{2}\right)\right), \tag{3.14}$$

where $\mu_{y_i|z_i}$ and $\boldsymbol{\Sigma}_{y|z}$ are as in Theroem 3.1.1. The expected probability of error given by (3.14) is not tractable for an analysis carried to characterize the optimal linear transform $\mathbf{T}$ that minimizes it. This stems from the fact that, $\mathbf{T}$ is a function of the statistics of the system, i.e., although it is easy to find a linear transformation minimizing the conditional probability of error (since Q-function is monotonic in its argument) this linear transform (which is a function of realizations of $\mathbf{x}_0$ and $\mathbf{x}_1$) may fail to minimize conditional probability of error for other realizations, so to characterize the

expected probability of error minimizing transform (which, then, would be a function of statistics of the system) we must evaluate the expectation in (3.14). However, the result of the expectation operation, i.e., the $m \times m$-fold integration in (3.14) is not given in terms of standard analytical functions. Therefore, we continue our analysis by characterizing linear operator(s) that minimize *a tight upper bound* on the expected probability of error defined by (3.14).

By virtue of the above discussion on the tractability of the analysis of linear transforms minimizing P_e, we follow the following approach; we first bound the conditional probability of error for any given pair of $\{\mathbf{z}_0\,,\,\mathbf{z}_1\}$ from above and make use of the fact that expected value of this upper bound is an upper bound on the expected probability of error (since, by definition, $\Pr[\text{error}\,|\,\{\mathbf{z}_0\,,\,\mathbf{z}_1\}] \geq 0$). The use of an *upper* bound clearly makes sense since we aim to *minimize* our objective function, i.e., P_e. The upper bound on $\Pr[\text{error}\,|\,\{\mathbf{z}_0\,,\,\mathbf{z}_1\}]$ that we use is the *Chernoff bound* on the Q-function (see *Basic Inequality* in [16]). This bound is an exponentially decaying and hence a tight upper bound on Q-function. The expected Chernoff bound, which replaces the primary objective function P_e in the design of optimal linear transfom $\mathbf{T}$ due to its analytical tractability and sufficient tightness, is derived in the following proposition.

Proposition 3.2.1 *The Chernoff bound on* $\Pr[\textit{error}\,|\,\{\mathbf{z}_0\,,\,\mathbf{z}_1\}]$ *is given by*

$$\Pr[\textit{error}\,|\,\{\mathbf{z}_0\,,\,\mathbf{z}_1\}] \leq \frac{1}{2}\exp\left(-\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0}-\mu_{y_1|z_1}\right)\|^2}{8}\right). \tag{3.15}$$

Also, the corresponding expected Chernoff bound on P_e *is*

$$P_e \leq \frac{1}{2}\left\{\det\left(\mathbf{I}_m+\frac{\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}}{2}\right)\right\}^{-\frac{1}{2}}. \tag{3.16}$$

Proof:

See Appendix B. ∎

The bound on expected (unconditional) probability of error of the MAP detector, given by (3.16) is the objective function we aim to minimize for this section. The minimization is carried over a class of linear transformations that posses certain properties imposed by the physical structure of the system analyzed. The obvious one of these properties is the dimensions of the transformations, and the other one is the constraint on its rank. The rank constraint is set to ensure that the partial information shared between the two sides of communication is not below or above a certain level, i.e., the space where the original signals are projected have a definite structure.

In the light of Proposition 3.2.1 and the discussion above we are ready to define the optimal linear transformation in the sense of probability of error bound minimization.

Definition 3.2.2 *The "probability of error bound minimizing transform* $\mathbf{T}^*$*" is defined as*

$$\mathbf{T}^* \stackrel{\triangle}{=} \underset{\substack{\mathbf{T}\in\mathbb{R}^{m\times n}\\ r(\mathbf{T})=m}}{\operatorname{argmax}} \quad \det\left(\mathbf{I}_m + \frac{\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}}{2}\right) \tag{3.17}$$

This definition of $\mathbf{T}^*$ not only defines the optimal $\mathbf{T}$ (in a probability of error bound sense) but also constructs our main problem in this section. More concretely, given the statistical structure of $\mathbf{x}_i, i \in \{0,1\}$ and $\mathbf{e}$ (all having zero mean normal distributions) we show how to construct $\mathbf{T}^*$ given by Definition 3.2.2.

The problem induced by the construction of the optimal transform $\mathbf{T}^*$ is characterized fully by the covariance matrices $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_e$, since the objective function to be maximized given in (3.17) is a function of $\mathbf{T}, \boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_e$. Below, we solve this problem, i.e., the construction of $\mathbf{T}^*$ for any given pair of matrices $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_e$ that are positive definite (as assumed in Sec. 2.2). Also special cases are studied to provide comprehension. We start with the following reduction of the original problem.

Proposition 3.2.2 *Let $\mathcal{S}_{\mathbf{T}}$ be the set of $m \times n$ rank m matrices and $\mathcal{S}_{\mathbf{M}}$ be the set of $m \times n$ orthonormal matrices, i.e., $\mathcal{S}_{\mathbf{T}} = \{\mathbf{T} \in \mathbb{R}^{m\times n} \,|\, r(\mathbf{T}) = m\}$ and $\mathcal{S}_{\mathbf{M}} = \left\{\mathbf{M} \in \mathbb{R}^{m\times n} \,|\, \mathbf{M}^T\mathbf{M} = \mathbf{I}_m\right\}$. Suppose there exists $\mathbf{M}^* \in \mathcal{S}_{\mathbf{M}}$ that satisfies*

$$\mathbf{M}^* = \underset{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}}{\operatorname{argmax}} \quad \prod_{i=1}^{m} \quad 1 + \frac{1}{\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_{\mathbf{P}}\mathbf{M}\right)}, \tag{3.18}$$

where $\hat{\mathbf{\Lambda}}_{\mathbf{P}} = \mathbf{I}_n - \mathbf{\Lambda}_{\mathbf{P}}^{-1}$; $\mathbf{\Lambda}_{\mathbf{P}}$ denoting the diagonal matrix of eigenvalues of $\mathbf{P} = \mathbf{\Lambda}^{-1}\mathbf{F}^T(\mathbf{\Sigma}_x + \mathbf{\Sigma}_e)\mathbf{F}\mathbf{\Lambda}^{-1}$ for which $\mathbf{F}$ and $\mathbf{\Lambda}$ denote matrix of eigenvectors and diagonal matrix of eigenvalues of $\mathbf{\Sigma}_x$, respectively. Then, there also exist $\mathbf{T} \in \mathcal{S}_{\mathbf{T}}$ satisfying (3.17), where cardinality of $\{\mathbf{T} \in \mathcal{S}_{\mathbf{T}} | \mathbf{T} = \mathbf{T}^*\}$ is that of the continuum.

Proof:

See Appendix C. ∎

Proposition 3.2.2 allows us to deduce the existence of $\mathbf{T}^*$ with the sufficiency of the existence of $\mathbf{M}^*$. Then, in order to construct an optimal linear transformation to extract partial information, which is the main goal of this section, we first need to show the existence of $\mathbf{M}^*$ and then construct $\mathbf{T}^*$ from the construction of $\mathbf{M}^*$ that is the solution for the reduced problem (3.18). In the following proposition we handle the first one of these tasks, i.e., the constructive solution to the existence of $\mathbf{M}^*$.

Proposition 3.2.3 *A set of solutions for (3.18) is given by*

$$\mathcal{M} = \left\{ \mathbf{M} \in \mathcal{S}_{\mathbf{M}} \;\middle|\; \mathbf{M} = \begin{bmatrix} \mathbf{\Gamma}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix},\ \mathbf{\Gamma}_m \in \mathbb{R}^{m\times m} \text{ is a unitary matrix} \right\} \tag{3.19}$$

Moreover,

$$\max_{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}} \quad \prod_{i=1}^{m} \quad 1 + \frac{1}{\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_{\mathbf{P}}\mathbf{M}\right)} \quad = \quad \prod_{i=1}^{m} \quad 1 + \frac{1}{\lambda_i\left(\hat{\mathbf{\Lambda}}_{\mathbf{P}}\right)}. \tag{3.20}$$

Proof:

See Appendix D ∎

Having shown the existence of optimal linear transform $\mathbf{T}^*$ we are now ready to construct the optimal transform following the construction of the set $\mathcal{M}$ of $\mathbf{M}^*$ provided in Proposition 3.2.3 and the construction of the set of $\mathbf{T}^*$ for a given element of $\mathcal{M}$ (which was provided in Appendix C). The following theorem, which is a combination of the results in Proposition 3.2.2 and those in Proposition 3.2.3, constructs set of optimal linear transforms in the sense of minimum Chernoff bound on the expected probability of error, in a self-contained way.

Theorem 3.2.1 *A set of optimal linear transforms, in the sense of expected Chernoff bound on the probability of error P_e, for communication with partial information in Gaussian setup is given by*

$$\mathcal{T} = \left\{ \mathbf{T} \in \mathcal{S}_{\mathbf{T}} \mid \mathbf{T} = \mathbf{U_K} \mathbf{\Lambda_K} \mathbf{M}^T \mathbf{U_P}^T \mathbf{\Lambda}^{-1} \mathbf{F}^T;\ \mathbf{U_K} \text{ is unitary, } \mathbf{\Lambda_K} \text{ is diagonal, } \mathbf{M} \in \mathcal{M} \right\}, \tag{3.21}$$

where $\mathcal{S}_{\mathbf{T}} = \{\mathbf{T} \in \mathbb{R}^{m \times n} \mid r(\mathbf{T}) = m\}$, $\mathcal{M}$ is given by Proposition 3.2.3 and $\mathbf{F}$, $\mathbf{\Lambda}$ and $\mathbf{U_P}$ denote matrix of eigenvectors and diagonal matrix of eigenvalues of $\mathbf{\Sigma}_x$ and the matrix of eigenvectors of $\mathbf{P} = \mathbf{\Lambda}^{-1} \mathbf{F}^T (\mathbf{\Sigma}_x + \mathbf{\Sigma}_e) \mathbf{F} \mathbf{\Lambda}^{-1}$, respectively. Moreover, the cardinality of $\mathcal{T}$ is equal to that of the continuum.

Proof:

By Proposition 3.2.2 we know that $\mathcal{T} \neq \emptyset$, also by Lemma C.0.2 in Appendix C, we know for given $\mathbf{M}^*$, i.e. $\mathbf{M}$ satisfying (3.18), $\mathbf{T} = \mathbf{U_K} \mathbf{\Lambda_K} \mathbf{M^*}^T \mathbf{U_P}^T \mathbf{\Lambda}^{-1} \mathbf{F}^T$ satisfies (3.17), i.e., $\mathbf{T} = \mathbf{T}^*$. Moreover, a set of $\mathbf{M}$ satisfying (3.18), namely $\mathcal{M}$, is given by Proposition 3.2.3. This clearly implies that $\mathcal{T}$, induced by $\mathcal{M}$, is a set of optimal linear transforms, in the sense of expected Chernoff bound on the probability of error P_e, for communication with partial information in Gaussian setup. Finally, since by Proposition 3.2.2 we know that $\mathcal{M} \neq \emptyset$ implies that cardinality of $\mathcal{T}$ is that of the continuum, we are done by Proposition 3.2.3. ∎

Theorem 3.2.1 gives us a complete characterization for a (large) set of optimal solutions for the problem given by Definition 3.2.2. In the following remark, we study the behavior of optimal transforms given by this solution compared to arbitrary transforms, which are elements of $\mathcal{S}_{\mathbf{T}}$. Also we study the behavior with respect to changing system parameters and provide further analytical results for a special case.

Remark 3.2.1 *The provable optimality of linear transforms is analyzed in this remark. We first illustrate this optimality.*

- Optimality of T^*: *Theorem 3.2.1 gives a set of optimal linear transforms, however does not address the "denseness" of $\mathcal{T}$ in $\mathcal{S}_{\mathbf{T}}$, i.e. the question of "is it easy to find an optimal transform in $\mathcal{S}_{\mathbf{T}}$ randomly, and how much is the performance of transforms in $\mathcal{S}_{\mathbf{T}} \backslash \mathcal{T}$ separated from that of optimal transforms?". The computational results in Figure 3.1 give us a basis to discuss this issue. In Figure 3.1 the*

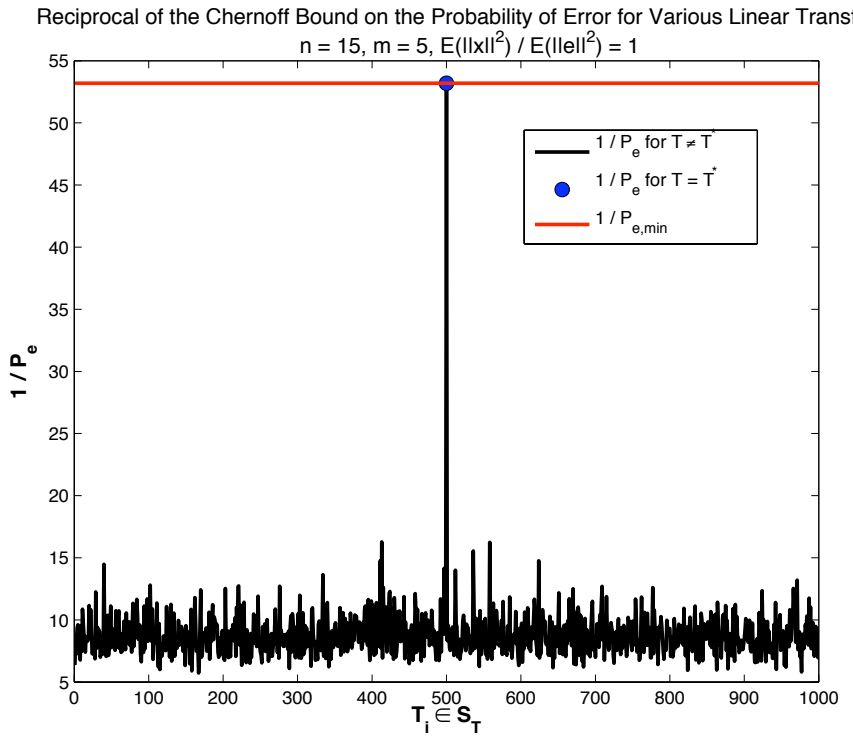

Figure 3.1. Performance of $\mathbf{T}^*$ compared to arbitrary $\mathbf{T} \in \mathcal{S}_{\mathbf{T}}$, P_e indicates Chernoff bound on expected probability of error here.

simulations are performed with $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_e$ having uniformly distributed eigenval-

ues, and the result is given using the reciprocal of the Chernoff bound on P_e to improve visibility. This figure shows two simple results. One result is that it is not "easy" to guess an element of $\mathcal{T}$ randomly (we actually simulated over much larger number of trials, however give here the result for a set of 1000 *trials for illustrative purposes). This is clear by observing that none of the transforms chosen randomly from $\mathcal{S}_{\mathbf{T}}$ achieves the optimal value calculated from (3.20) in Proposition 3.2.3, except $\mathbf{T}^*$ constructed by (3.21) and indicated as the transform in the middle of set of transforms, i.e. $\mathbf{T}_{500}$. Also, the minimum value of the bound on P_e achieved by arbitrary choices is not even close to that achieved by $\mathbf{T}^*$, it is around* 4 *times larger than the minimum bound on P_e (moreover, the difference in performances is much larger for other choices of n, m and $E(\|\mathbf{x}\|^2)/E(\|\mathbf{e}\|^2)$, but again we hold this difference in a visible range). By these two results, it is possible to claim that $\mathcal{T}$ is not "dense" in $\mathcal{S}_{\mathbf{T}}$, i.e. construction for optimality gives a distinguished performance in the sense of Chernoff bound on P_e.*

- P_e vs. $\mathrm{E}(\|\mathbf{x}\|^2)/\mathrm{E}(\|\mathbf{e}\|^2)$*: In this part we observe the effect of SNR defined by $SNR = E(\|\mathbf{x}\|^2)/E(\|\mathbf{e}\|^2)$ on the optimality of $\mathbf{T}^*$. Figure 3.2 is given to discuss this effect. Similar to Figure 3.1, the simulations in Figure 3.2 are performed with $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_e$ having uniformly distributed eigenvalues, where the effect of $SNR = tr(\mathbf{\Sigma}_x)/tr(\mathbf{\Sigma}_e)$, for $tr(.)$ representing the trace operator, is taken into consideration. The result is clear and expected; as SNR increases the performance at optimality is improved, since for higher SNR it is possible to discriminate $\mathbf{z}_0$ from $\mathbf{z}_1$ better, which then improves detection if we consider ML detection rule given by (3.9) and $\mu_{y_i|z_i} = E(\mathbf{y}_i \,|\, \mathbf{z}_i)\big|_{\mathbf{y}_i=\mathbf{x}_i+\mathbf{e}} = \mathbf{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)^{-1}\mathbf{z}_i$.*
- P_e vs. m*: In this case, we study the effects of the amount of partial information shared by the detector side on the bound on the expected performance of the detector. In other words, we observe the effect of the information available to the detector to construct its "codebook" for the binary data. This case is studied for $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_e$ having uniformly distributed eigenvalues and $SNR = 1$. For $n = 50$, we construct $\mathbf{T}^*$ for particular values of m and evaluate its performance in the sense of expected Chernoff bound on P_e. Results are plotted in Figure 3.3. Again the result is as expected; for increasing amount of partial information the capability of the detector to make a decision between $\mathbf{z}_0$ and $\mathbf{z}_1$ increases, so we*

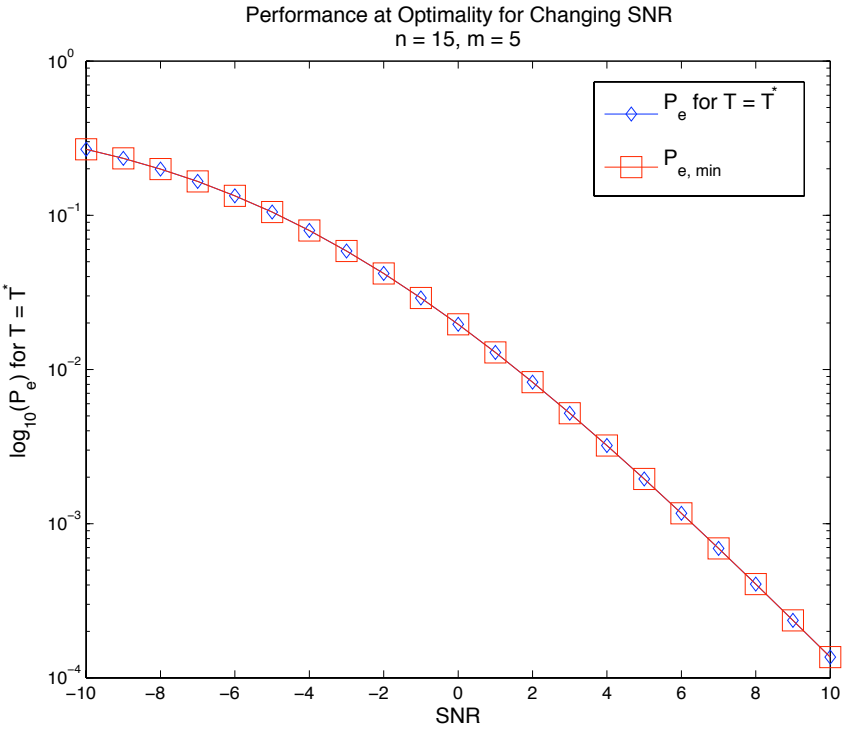

Figure 3.2. Performance of $\mathbf{T}^*$ for changing SNR, P_e indicates Chernoff bound on expected probability of error here. Also, SNR is in dB.

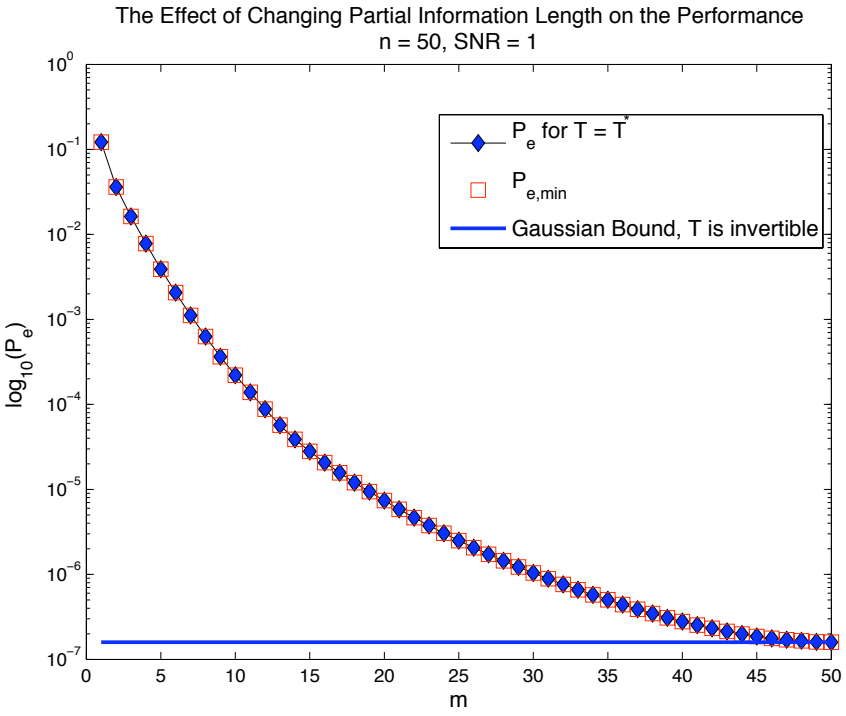

Figure 3.3. Performance of $\mathbf{T}^*$ for changing partial information length, P_e indicates Chernoff bound on expected probability of error here.

observe a decreasing optimal bound, which is achieved by $\mathbf{T}^*$*. An important observation is that as* m *increases to* n*, the performance at optimality converges to that for* $m = n$*, which is the Gaussian bound (the case for* $\mathbf{T}$ *is invertible) given by* $\frac{1}{2}\left\{\det\left(\mathbf{I}_n + \frac{\mathbf{\Sigma}_x\mathbf{\Sigma}_e^{-1}}{2}\right)\right\}^{-\frac{1}{2}}$*, derived from (3.16).*

- P_e vs. n*: In this part we study the effect of changes in signal length on the performance of* $\mathbf{T}^*$*. The simulations, which were carried for various* $\mathbf{\Sigma}_x$ *and* $\mathbf{\Sigma}_e$ *all having uniformly distributed eigenvalues, are depicted in Figure 3.4. This time,*

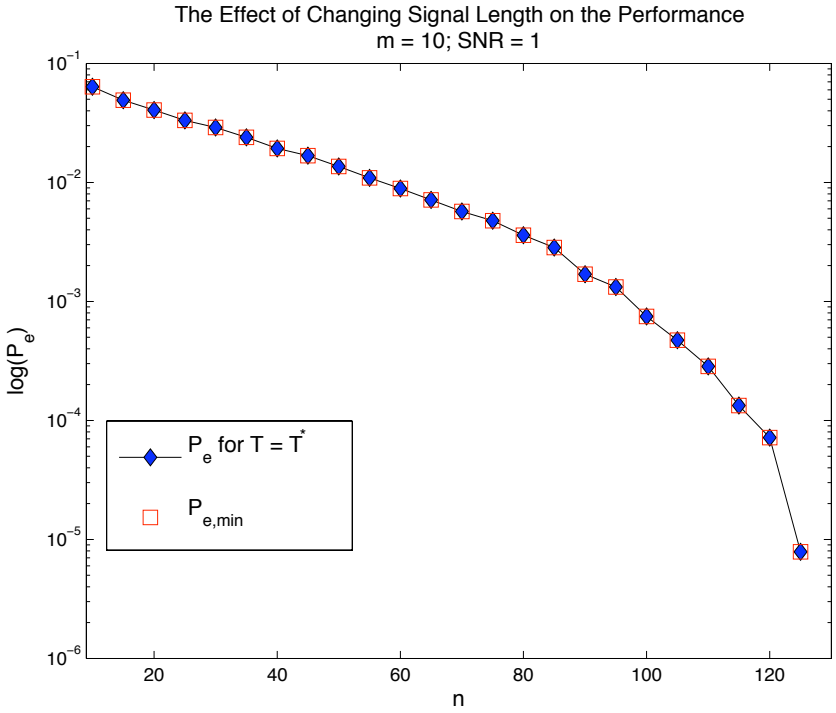

Figure 3.4. Performance of $\mathbf{T}^*$ for changing signal length, P_e indicates Chernoff bound on expected probability of error here.

the results might seem to be counter-intiutive if one forgets the task of the detector, i.e. binary detection using partial information*. The crucial point is that since* m *is constant, i.e. the dimension of the space of partial information is constant, as* n *increases we get more degrees of freedom to construct the optimal transform* $\mathbf{T}^*$ *(i.e. the cardinality of the set of eigenvalues of* $\mathbf{P}$ *increases and hence we get a possibly higher optimality given by (3.20) that* $\mathbf{M}^*$ *achieves, hence optimality of corresponding* $\mathbf{T}^*$ *is improved). One can conclude that, since we increase the redundancy in the signal representing binary information, the performance of decoding is improved, which is similar to an argument for error-correction coding.*

- The i.i.d. Case*: Now consider the case where we have* $\mathbf{\Sigma}_x = \sigma_x^2 \mathbf{I}_n$ *and* $\mathbf{\Sigma}_e = \sigma_e^2 \mathbf{I}_n$*. Then, the objective function in (3.17) gives*

$$\begin{aligned} \det\left(\mathbf{\Sigma}_{zAB}\right) &= \det\left(\mathbf{I}_m + \frac{\mathbf{V}_\mathbf{T}^T}{2}\left(\left(1+\frac{\sigma_e^2}{\sigma_x^2}\right)\mathbf{I}_n - \mathbf{V}_\mathbf{T}\mathbf{V}_\mathbf{T}^T\right)^{-1}\mathbf{V}_\mathbf{T}\right) && (3.22)\\ &= \det\left(\mathbf{I}_m + \frac{1}{2}\frac{\sigma_x^2}{\sigma_x^2+\sigma_e^2}\mathbf{V}_\mathbf{T}^T\left(\mathbf{I}_n + \frac{\sigma_x^2}{\sigma_e^2}\mathbf{V}_\mathbf{T}\mathbf{V}_\mathbf{T}^T\right)\mathbf{V}_\mathbf{T}\right) && (3.23)\\ &= \left(1+\frac{\sigma_x^2}{2\sigma_e^2}\right)^m && (3.24) \end{aligned}$$

where $\mathbf{\Sigma}_{zAB} \stackrel{\triangle}{=} \mathbf{I}_m + \frac{\mathbf{\Sigma}_z\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{2}$ *for which* $\mathbf{A} = \mathbf{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)^{-1}$, $\mathbf{B} = \mathbf{\Sigma}_{y|z}$, $\mathbf{\Sigma}_z = \mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T$ *and* $\mathbf{V}_\mathbf{T}$ *is the orthogonal matrix of right singular vectors of* $\mathbf{T}$*. Here, the first equality follows from properties of determinant (and taking into parentheses of invertible product of unitary matrix of left singular vectors and diagonal matrix of singular values of* $\mathbf{T}$*), the second one follows from the fact that* $\left(\mathbf{I}_n - a\mathbf{V}_\mathbf{T}\mathbf{V}_\mathbf{T}^T\right)^{-1} = \mathbf{I}_n - \frac{a}{a-1}\mathbf{V}_\mathbf{T}\mathbf{V}_\mathbf{T}^T$ *(you can verify by direct calculation) for* $a \neq 1$ *and it is simple to derive (3.24) from the second equality. Also, this direct calculation completely matches with the Gaussian bound equal to* $\frac{1}{2}\left\{\det\left(\mathbf{I}_n + \frac{\mathbf{\Sigma}_x\mathbf{\Sigma}_e^{-1}}{2}\right)\right\}^{-\frac{1}{2}}$ *for the case* $m = n$*.*

Here, by (3.24) we can conclude that there is nothing to optimize with respect to $\mathbf{T}$ *for given* m*, i.e.,* $\mathcal{T} = \mathcal{S}_\mathbf{T}$*. This is an expected result, since there is nothing "biased" by the statistical structure. However, there is one important point; as the size of partial information increases the optimality point is improved. This is again an expected result, because as* m *increases the capability of the detector to discriminate in a higher dimensional space of partial information increases.*

Moreover, for a case where the ratio between n *and* m *is kept fixed, i.e.,* $m/n = k$*, we have the following bound on the error exponent for our system (by (3.16))*

$$\lim_{n\to\infty} -\frac{1}{n}\log P_e \quad \geq \quad 2k\log\left(1+\frac{\sigma_x^2}{2\sigma_e^2}\right). \qquad (3.25)$$

In this section, we have fully characterized the structure of optimal linear transforms in the sense of expected Chernoff bound on the probability of error and discussed its behavior for various cases. Now we analyze communication with partial information

(with no jammer) under a different criterion for optimality.

3.3. Optimal Linear Operators In The Probabilistic Sense

The criterion induced by the concept of expected (unconditional) probability of error, which was studied in Sec. 3.2, is not the only or the most important criterion for the system given by Figure 2.1 for $P = 0$. For instance, instead of minimizing the expected probability of error, one may aim to solve the problem of minimizing the probability of the cases for which the conditional probability of detection error exceeds a threshold, which, for instance, is more meaningful if we consider the "high" error cases as undesired and "low" error cases as acceptable. In this section, we analyze the system described by Figure 2.1 for $P = 0$ exactly under this criterion, i.e. we aim to characterize optimal linear transforms which minimize the probability of the cases for which the conditional probability of detection error is above a pre-determined threshold value. This criterion is set to design the optimal linear transform in order to exclude cases of high error completely from the system, therefore it imposes a "harder" limitation on the performance. Similar to the case with expected probability of error criterion, we use binary MAP detection for given side information, i.e., the detector operates in the same way as it does in previous section, however the criterion is different from that in previous section.

For the criterion introduce above, we first define the "failure probability for the constant α".

Definition 3.3.1 *"The failure probability for the constant α", denoted by P_α for detection with partial information in additive Gaussian channel is*

$$P_\alpha = \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\Pr\left[error \mid \{\mathbf{z}_0,\mathbf{z}_1\}\right] > \alpha\right], \tag{3.26}$$

where $\Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}[.]$ denotes the probability evaluated with respect to the joint distribution of $\mathbf{z}_0$ and $\mathbf{z}_1$ and $\Pr\left[error \mid \{\mathbf{z}_0,\mathbf{z}_1\}\right]$ is given by (3.10).

This expression is clearly trivial for $\alpha \geq 1/2$ with $P_\alpha = 0$. So we assume $\alpha \leq 1/2$. Now, if we consider $\Pr[\text{error} \mid \{\mathbf{z}_0, \mathbf{z}_1\}]$ as a function of $\{\mathbf{z}_0, \mathbf{z}_1\}$, which then becomes a random variable, and denote the cumulative distribution function of $\Pr[\text{error} \mid \{\mathbf{z}_0, \mathbf{z}_1\}]$ evaluted at α by $F_{P(z_0,z_1)}(\alpha)$ the expression for P_α given by (3.26) satisfies $P_\alpha = 1 - F_{P(z_0,z_1)}(\alpha)$ (assuming α is a continuity point of $F_{P(z_0,z_1)}$), this is why we use the phrase "the cumulative distribution function of the conditional probability of error" in Sec. 2.2 to identify the problem studied in this section.

Here, if we consider Definition 3.3.1 together with (3.10), we get

$$P_\alpha = 1 - \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\mathrm{Q}\left(\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)\|}{2}\right) \leq \alpha\right]. \tag{3.27}$$

So, in order to minimize the probability of exceeding the conditional probability of error threshold α, i.e. P_α, we aim to maximize the cumulative distribution function of $\Pr[\text{error} \mid \{\mathbf{z}_0, \mathbf{z}_1\}]$ evaluted at α, i.e. $F_{P(z_0,z_1)}(\alpha)$. Now, we can further simplify $F_{P(z_0,z_1)}(\alpha)$ in the following way

$$\begin{aligned}
F_{P(z_0,z_1)}(\alpha) &= \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\mathrm{Q}\left(\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)\|}{2}\right) \leq \alpha\right] && (3.28)\\
&= \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)\|^2 \geq 4\left[\mathrm{Q}^{-1}(\alpha)\right]^2\right] && (3.29)\\
&= \Pr_{\gamma}\left[\gamma^T\mathbf{G}\gamma \geq \delta\right] && (3.30)
\end{aligned}$$

where $\gamma \triangleq \mathbf{z}_0 - \mathbf{z}_1$, $\mathbf{G} \triangleq \mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}$ for $\mathbf{A} \triangleq \boldsymbol{\Sigma}_x\mathbf{T}^T(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T)^{-1}$ and $\mathbf{B} \triangleq \boldsymbol{\Sigma}_{y\,|\,z}$ and $\delta = 4\left[\mathrm{Q}^{-1}(\alpha)\right]^2$. Here, (3.28) is equal to (3.29) since Q(.) is an invertible decreasing function of its argument and (3.29) is equal to (3.30) by Theorem 3.1.1. So, our objective function boils down to one minus the cumulative distribution of the sum of squares of (possibly) correlated normal random variables. We discuss the analytical tractability of the cumulative function of this random variable, which is a function of γ, in Remark 3.3.1.

Remark 3.3.1 *First of all, since* $\mathbf{G}$ *is symmetric, by spectral decomposition (see [15])*

we have $\mathbf{G} = \mathbf{U_G}\mathbf{\Lambda_G^2}\mathbf{U_G^T}$*, where* $\mathbf{U_G}, \mathbf{\Lambda_G} \in \mathbb{R}^{m\times m}$ *are unitary and invertible diagonal (since* $\mathbf{G} > 0$*, see Appendix B). Now let* $\beta \triangleq \mathbf{\Lambda_G}\mathbf{U_G^T}\gamma$*. So we have* $\Pr_\gamma\left[\gamma^T\mathbf{G}\gamma \geq \delta\right] = \Pr_\beta\left[\beta^T\beta \geq \delta\right]$*, yielding by (3.30)*

$$F_{P(z_0,z_1)}(\alpha) = 1 - \Pr_\beta\left[\beta^T\beta < \delta\right]. \tag{3.31}$$

So our aim becomes minimizing the objective function $\Pr_\beta\left[\beta^T\beta < \delta\right]$*. Here, by the definition of* β*, we know* $\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\beta)$*, where* $\mathbf{\Sigma}_\beta = \mathbf{\Lambda_G}\mathbf{U_G}\mathbf{\Sigma}_\gamma\mathbf{U_G^T}\mathbf{\Lambda_G}$ *(see [17]), which is clearly a symmetric covariance matrix. Then, by spectral decomposition if we let* $\mathbf{\Sigma}_\beta = \mathbf{U}_\beta\mathbf{\Lambda}_\beta\mathbf{U}_\beta^T$*, we get*

$$\begin{aligned}
\Pr_\beta\left[\beta^T\beta < \delta\right] &= \Pr_\beta\left[tr\left(\beta\beta^T\right) < \delta\right] &(3.32)\\
&= \Pr_\beta\left[tr\left(\mathbf{U}_\beta^T\beta\beta^T\mathbf{U}_\beta\right) < \delta\right] &(3.33)\\
&= \Pr_\beta\left[\|\rho\|^2 < \delta\right], &(3.34)
\end{aligned}$$

where (3.33) follows from the properties of trace operator and $\rho \triangleq \mathbf{U}_\beta^T\beta$ *is a zero mean normal random vector with a covariance of* $\mathbf{\Sigma}_\rho = \mathbf{\Lambda}_\beta$ *(see [17]), i.e. it is a vector composed of uncorrelated (hence independent, by normallity) but not identically distributed normal random variables. Then, our objective function reduces to the cumulative distribution function of the sum of* m *independent (since measurable functions of independent random variables are independent and polynomials are measurable) chi-squared random variables with non-identical densities. Since they are non-identically distributed,* $\sum_{i=1}^m \rho_i^2$ *is* not *a chi-squared random variable with* m *degrees of freedom.*

To see if we can derive the cumulative distribution for $\|\rho\|^2$*, we consider its characteristic function (since there is a one-to-one correspondence between characteristic functions and corresponding distribution functions, see [16]), which is given by the multiplication of the characteristic functions of independent components of* ρ *(see [16]).*

Then we can derive (see [17] for characteristic function of one component)

$$\phi_{\sum \rho_i^2} = \left(\prod_{k=1}^{m} \Delta_k^{1/2} \right) \exp\left(\frac{\sqrt{-1}}{2} \sum_{k=1}^{m} \arccos\left(\Delta_k\right) \right) \quad \textit{where} \quad \Delta_k = \left(1 + 4\mathbf{\Lambda}_{\beta;k,k}^2 t^2\right)^{-1/2}, \tag{3.35}$$

where $\mathbf{\Lambda}_{\beta;k,k}$ indicates the k^{th} diagonal entry of $\mathbf{\Lambda}_\beta$. The inverse Fourier transform of $\phi_{\sum \rho_i^2}$ corresponds to distribution function of $\|\rho\|^2$, for continuity points of the distribution function (see [16]), which is supposed to yield our objective function for α taken as a continuity point. This inversion seems impossible to tract in terms of known analytical functions, since $\mathbf{\Lambda}_{\beta;k,k}$ may change for various k, ttherefore we follow the approach of constructing a bound for P_α and analyzing this bound.

Because of the difficulty of the analytical tractability of P_α, and equivalently that of $\Pr_\beta\left[\beta^T\beta < \delta\right]$ for constructing optimal linear transforms, we set an upper bound on $\Pr_\beta\left[\beta^T\beta < \delta\right]$ that is derived in terms of the eigenvalues of $\mathbf{\Sigma}_\beta$ and analyze optimal transforms achieving the optimal value of this bound. Since we aim to minimize $\Pr_\beta\left[\beta^T\beta < \delta\right]$, which is equivalent to minimizing P_α, we set an *upper* bound to be minimized, as in Sec. 3.2. The following proposition gives this bound on $\Pr_\beta\left[\beta^T\beta < \delta\right]$.

Proposition 3.3.1 *The bound on P_α is given by*

$$P_\alpha \leq \mathcal{P}\left(\frac{m}{2}, \frac{\delta}{4\lambda_{min}\left(\mathbf{T}\mathbf{\Sigma}_x \mathbf{\Sigma}_{y\,|\,z}^{-1} \mathbf{\Sigma}_x \mathbf{T}^T \left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)^{-1} \right)} \right), \tag{3.36}$$

where $\mathcal{P}\left(.,.\right)$ is the regularized gamma function.

Proof:

See Appendix E ∎

The above proposition allows us to consider a cost function that is analytically tractable in terms of $\mathbf{T}$, which is the only parameter of system we have control on. We aim to minimize the cost given by right hand side of (3.36) over the set of full-rank

$m \times n$ real matrices, where again as in previous case we, we want to restrict the partial information shared between the two sides of communication and hence require $m < n$. On the other hand, we want that the partial information provided to the detector side is efficient, which then imposes the full-rank property for the linear transformation used to extract this information from the original signal.

Since we have the (tractable) cost function given by (3.36), we can define the optimal transform, using that the regularized gamma function is strictly monotonic in both of its arguments. Hence we have the following definition for the problem of concern.

Definition 3.3.2 *The "linear transform minimizing the bound on the cumulative distribution function of conditional probability of error" is defined as*

$$\mathbf{T}^* \stackrel{\triangle}{=} \underset{\substack{\mathbf{T}\in\mathbb{R}^{m\times n}\\ r(\mathbf{T})=m}}{\operatorname{argmax}} \quad \lambda_{min}\left(\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\right). \tag{3.37}$$

By Definition 3.3.2 we also provide the main problem of this section, i.e. characterizing the optimal linear transform given by (3.37). Similar to the previous on optimal linear transforms in expected sense, we follow a constructive approach to this problem. The construction of a set of optimal linear transforms is accomplished by first providing a reduction of the original problem, for which we have a corresponding set of optimal transforms in the sense of (3.37), and then show how we can construct the optimality "points" for this reduced problem. The results of this approach, which is similar to that in the previous section, are given in Theorem 3.3.1, where the above described steps are followed in its proof in Appendix F.

Theorem 3.3.1 *A set of optimal linear transforms, in the sense of the bound on the cumulative distribution function of the conditional probability of error in (3.37), is given by $\mathcal{T}$ in (3.21), i.e., if $\mathcal{T}$ is optimal in the sense of Definition 3.2.2 then it is also optimal in the sense of Definition 3.3.2, achieving the maximum value given by*

$\mathcal{P}\left(m/2, \delta/4\left(\frac{1}{\lambda_m(\hat{\mathbf{\Lambda}}_\mathbf{P})} - 1\right)\right)$, *where* $\lambda_m\left(\hat{\mathbf{\Lambda}}_\mathbf{P}\right)$ *indicates the* m^{th} *smallest eigenvalue of* $\hat{\mathbf{\Lambda}}_\mathbf{P}$ *given in Proposition 3.2.2.*

Proof:

See Appendix F. ∎

Remark 3.3.2 *The minimum value of* $\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)$ *can be achieved also by a linear transform with less number of rows, i.e. for smaller* m*. So a question arises: "Is it enough to use only* 1 *dimensional partial information to achieve bound minimizing optimality for the criterion of minimizing the bound on the cumulative distribution function of conditional probability of error?". The answer is "No", because if we have* 1 *dimensional partial information we achieve the bound* $\mathcal{P}\left(1/2, max\left\{\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\right\}\delta/2\right)$ *where* $\mathcal{P}\left(1/2, max\left\{\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\right\}\delta/2\right) > \mathcal{P}\left(m/2, max\left\{\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\right\}\delta/2\right)$*, and* $\mathcal{P}\left(m/2, max\left\{\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\right\}\delta/2\right)$ *indicates the bound achieved by* $\mathbf{T}^*$ *(for* $m > 1$*). Therefore, although we achieve the same cost function given by (3.37) for given* m*, the bound achieved on* P_α *decreases as* m *increases, i.e. providing more partial information decreases the bound on* P_α*, which is expected.*

As in the case of previous section, Theorem 3.3.1 gives us a complete characterization for a (large) set of optimal solutions for the problem given by Definition 3.3.2. In Remark 3.3.3, we study the optimal transforms in terms of the effects of the system parameters on performance and we also qualitatively observe the superiority of optimal transforms to arbitrary transforms in $\mathcal{S}_\mathbf{T}$.

Remark 3.3.3 *This remark studies the overall behavior of optimal transforms defined by (3.3.2).*

- Optimality of T^**: Similar to Theorem 3.2.1, Theorem 3.3.1 does not address the issue of the superiority of* $\mathbf{T}^*$ *to other full-rank linear transforms of size* $m \times n$*. Figure 3.5 addresses this issue qualitatively. Again, to improve visibility of the*

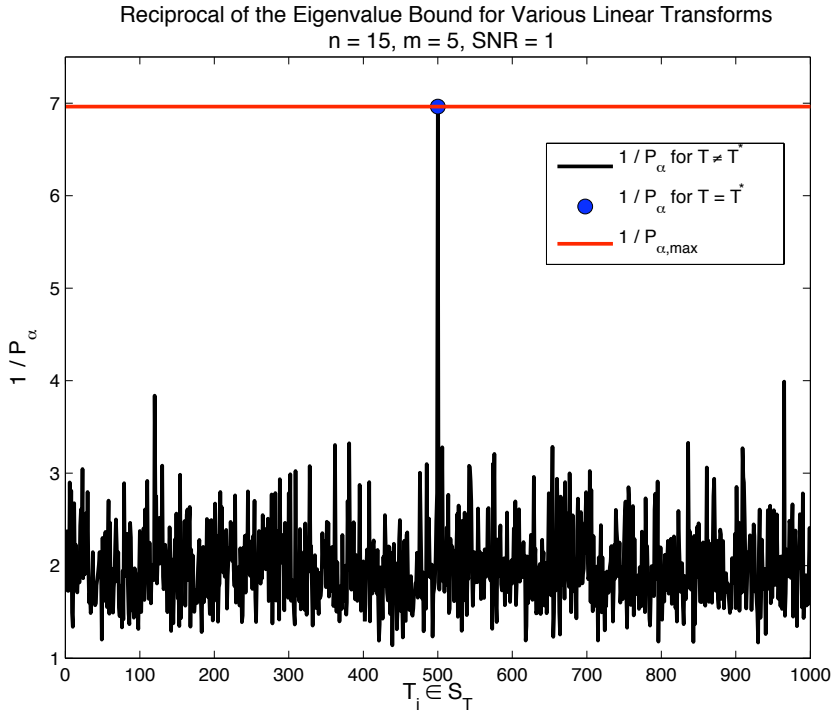

Figure 3.5. Performance of $\mathbf{T}^*$ compared to arbitrary $\mathbf{T} \in \mathcal{S}_{\mathbf{T}}$, P_α indicates eigenvalue bound on P_α.

difference, we give the results in terms of their reciprocals. The superiority of $\mathbf{T}^*$ *is clear from the figure, where the optimal bound achieved by* $\mathbf{T}^*$ *is more than* 3 *times smaller than that for other randomly chosen transforms in* $\mathcal{S}_{\mathbf{T}}$*. Also, as in the previous section, the performance of none of the random transforms is "close" to that of* $\mathbf{T}^*$*, allowing us to claim about the distinguished performance of* $\mathbf{T}^*$ *compared to randomly chosen transforms. Here, simulations are performed for* $\mathbf{\Sigma}_x$ *and* $\mathbf{\Sigma}_e$ *having uniformly distributed eigenvalues.*

- P_α vs. α*: In this part of the remark, we study the behavior of* P_α *with respect to* α*. This relation between* P_α *and* α *is easy to discuss about, since* $\Pr\left[error \mid \{\mathbf{z}_0, \mathbf{z}_1\}\right]$ *has a strictly increasing cumulative function between* 0 *and* 1*, which yields that the bound on* $1 - \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\Pr\left[error \mid \{\mathbf{z}_0, \mathbf{z}_1\}\right] < \alpha\right]$ *is strictly decreasing between* 1 *and* 0 *(if it is tight enough). Figure 3.6 illustrates the above inferences, where the simulations are performed for* $\mathbf{\Sigma}_x$ *and* $\mathbf{\Sigma}_e$ *with uniformly distributed eigenvalues.*

- P_α vs. SNR*: The effect of changing* SNR *is given by Figure 3.7, where, again, simulations are performed for* $\mathbf{\Sigma}_x$ *and* $\mathbf{\Sigma}_e$ *having uniformly distributed eigenval-*

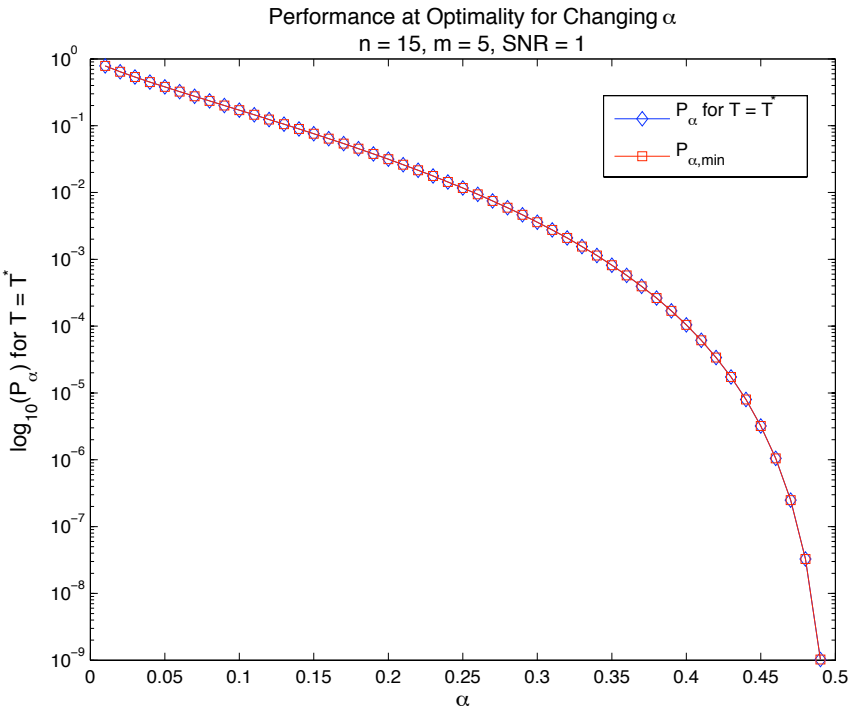

Figure 3.6. Performance of $\mathbf{T}^*$ for changing α, P_α indicates eigenvalue bound on P_α.

ues. The results are, simply, as expected; as SNR increases the conditional prob-

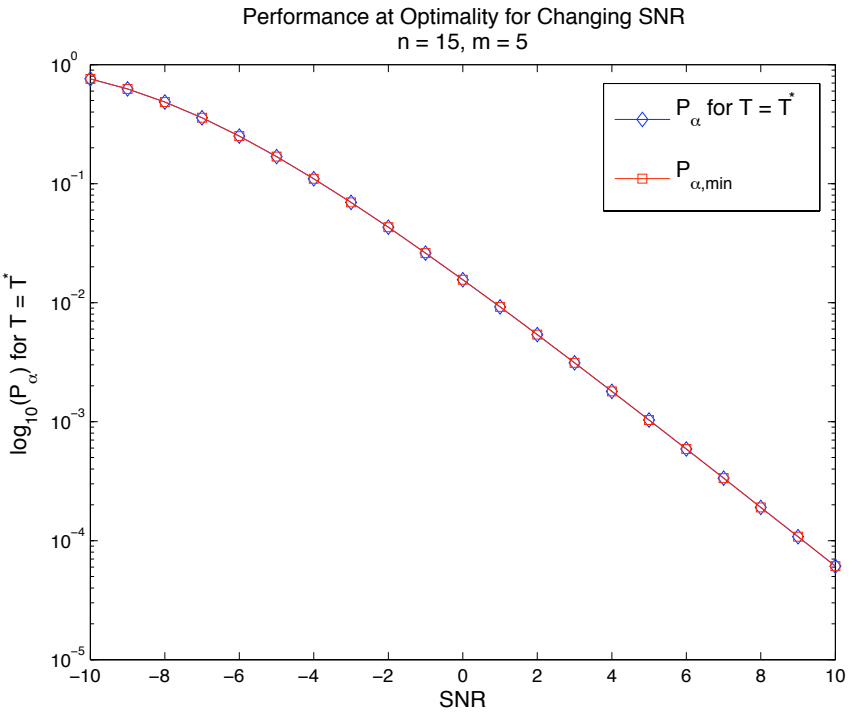

Figure 3.7. Performance of $\mathbf{T}^*$ for changing SNR, P_α indicates eigenvalue bound on P_α.

ability of error decreases and hence its cumulative distribution function increases for all α*. This yields the expected inverse relationship between the bound on* P_α*, which is* $1 - \Pr_{\{\mathbf{z}_0,\mathbf{z}_1\}}\left[\Pr\left[error \mid \{\mathbf{z}_0, \mathbf{z}_1\}\right] < \alpha\right]$*, and SNR.*

- P_α vs. m*: Here, we study the effect of partial information size* m *on the performance of* $\mathbf{T}^*$*. As in the expected error case, the increasing amount of partial information shared by the detector improves the performance at optimality since the capability of the detector to make a correct decision conditioned on partial information increases. The results are given in Figure 3.8. Also, as the partial information shared approaches "complete" information, i.e.* $m \longrightarrow n$*, the performance approaches to that of the case where* $\mathbf{T}^*$ *is invertible. Hence, the bound approaches to* $\mathcal{P}\left(n/2, \delta/4\left(\lambda_{min}\left(\mathbf{\Sigma}_x\mathbf{\Sigma}_e^{-1}\right)\right)\right)$*, which is derived by assuming invertibility of* $\mathbf{T}^*$ *in (3.37).*

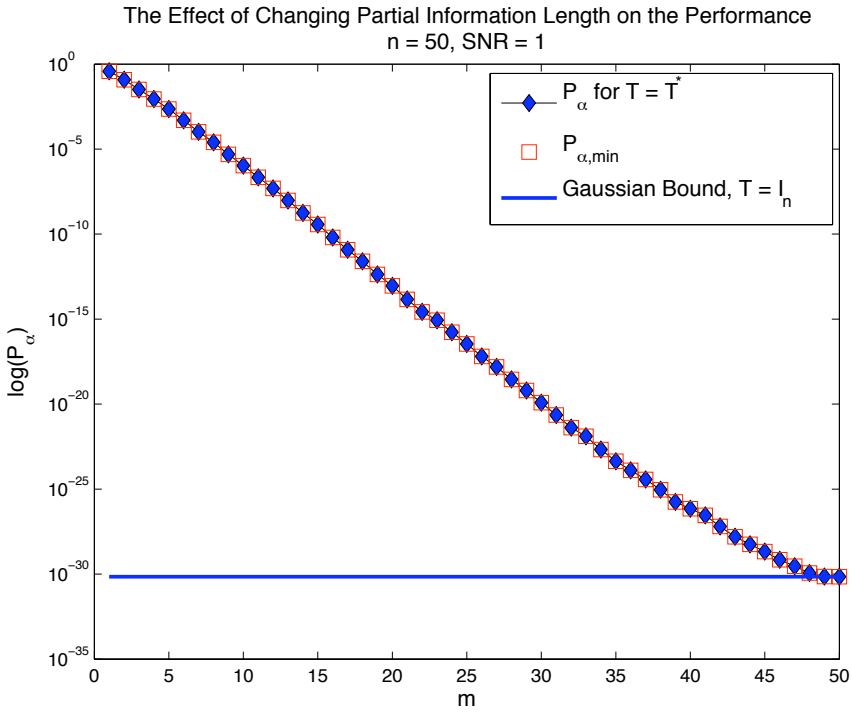

Figure 3.8. Performance of $\mathbf{T}^*$ for changing partial information size, P_α indicates eigenvalue bound on P_α.

- P_e vs. n*: As the last parameter of the system, we investigate the effect of transmitted signal size in this part. Similar to expected error case, we should not forget that the detector aims to make a decision regarding its index, i.e., aims to make a "binary decision". So, as the size of the originally transmitted signal increases*

the capability of the detector to distinguish between the corresponding partial information increases, and that is why the bound on the probability of the "failure" of the detector decreases. More concretely, as the size of the space characterized by transmitted signals increases, the possibility to find smaller eigenvalues for $\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}$ *increases, which in turn implies that the bound on the failure probability decreases. The results of simulations are given in Figure 3.9.*

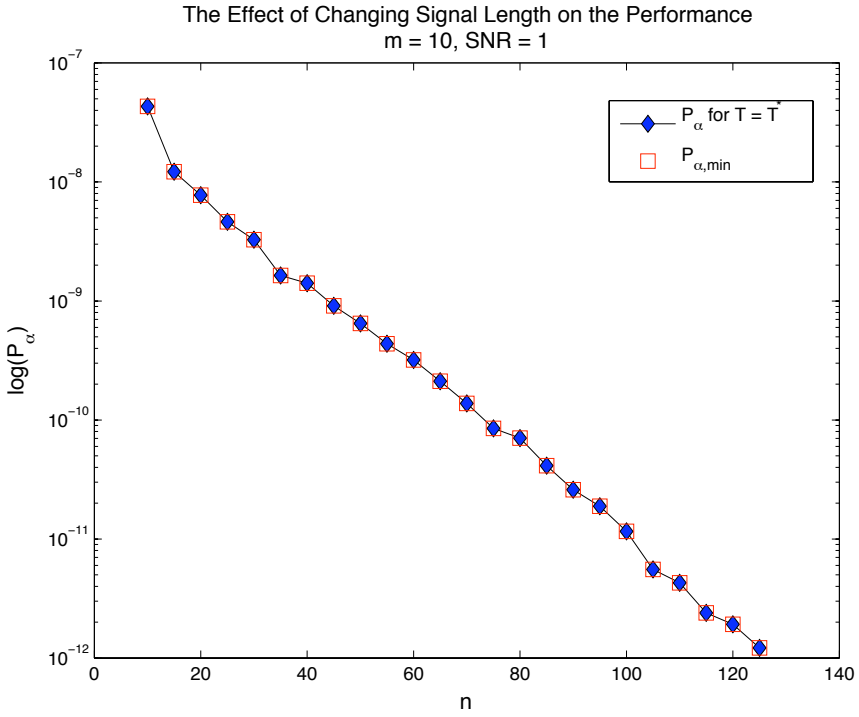

Figure 3.9. Performance of $\mathbf{T}^*$ for changing signal size, P_α indicates eigenvalue bound on P_α.

The above discussion in Remark 3.3.3 completes our study of optimal transforms in the sense of (3.37). Consequently, we have completed our study of the case with no jammer, i.e., the case for which $P = 0$, almost surely. Now, we proceed to formulate and analyze communication with partial information in the presence of an intelligent jammer, which has constraint capabilities.

4. DETECTION IN THE PRESENCE OF A JAMMER

In this section, the system in Figure 2.1 for $\|\mathbf{u}\|^2 \leq P$ a.s., where $P > 0$, is studied. In other words, we study communication with partial information in the presence of a jammer, where the jammer function is not necessarily 0 for a set of probability measure one, however is subject to the *peak power constraint* given by $\|\mathbf{u}\|^2 \leq P$ a.s.. As described in Sec. 2.2, jammer is assumed to be *all-knowing*, i.e., jammer is assumed to know the binary information i, the codewords of transmitter $\mathbf{x}_0$ and $\mathbf{x}_1$ and the linear transform $\mathbf{T}$. The only information jammer lacks is the realizations of noise $\mathbf{e}$, moreover we assume that $\mathbf{u} \stackrel{\triangle}{=} J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$ is independent of $\mathbf{e}$. One should note that, this characterization of $\mathbf{u}$ allows it to be a random function of its arguments, where we require $\mathbf{u}$ be a Borel measurable function of its arguments to make sense as a random variable.

If we consider the detector structure for this section, we should first note that it is not an MAP detector, which was the case for the additive Gaussian channel studied in previous sections. For the case including a jammer, the detector is assumed to be aware of the "existence" of the jammer, however it has no information on the "capabilities" of the jammer. This implies that (although the detector can solve for the optimal jammer behavior if it has "complete" characterization of the jammer) the detector is unaware of the statistics of the jammer, and hence cannot apply optimal MAP rule for binary detection. Instead, it applies a sub-optimal, however less parametric and hence less sensitive to unknown system parameters, method of *minimum Euclidean distance detection* in the subspace induced by the linear transform $\mathbf{T}$, i.e. in range space of $\mathbf{T}$. Then, the jammer aim is to worsen the performance of this detector by designing, potentially random, $\mathbf{u}$ *for each realization of transmission* in terms of the probability of binary detection. On the other hand, the linear transform designer aims to improve the performance of detection by designing the linear transform *once for all realizations of transmission* in terms of the probability of failure, as in Sec. 3.3.

After the qualitative characterization of our problem for this section given above, we start to formally state our problem by defining the detector operation.

Definition 4.0.3 *The detection rule for* minimum Euclidean distance detection *in the range space of* $\mathbf{T}$ *for equal priors on* $i = 0$ *and* $i = 1$ *is given by*

$$\|\mathbf{Ty} - \mathbf{z}_0\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \|\mathbf{Ty} - \mathbf{z}_1\|^2, \tag{4.1}$$

where $\mathbf{y}$ *is the observation at the output of the channel, and* H_i *is the hypothesis of that* $\mathbf{x}_i$ *is transmitted and* $\{\mathbf{z}_0, \mathbf{z}_1\}$ *are given.*

The above definition of the detector operation, by which the "error event" makes sense, allows us to define our main objective function for this section, namely "the worst-case failure probability".

Definition 4.0.4 *"The worst-case failure probability", denoted by* P_{α,J^*}*, is given by*

$$P_{\alpha,J^*} = \Pr{}_{\{\mathbf{z}_0,\mathbf{z}_1\}} \left[\Pr\left[error|\{\mathbf{z}_0, \mathbf{z}_1\}\right] \Big|_{\mathbf{u}=\mathbf{u}^*} > \alpha \right] \tag{4.2}$$

where the "worst-case jammer function", $\mathbf{u}^*$*, which is required to be a Borel measurable function of* $i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T}$ *is defined by*

$$\mathbf{u}^* \overset{\triangle}{=} \underset{\|\mathbf{u}\|^2 \le P,\ a.s.}{\operatorname{argmax}} \quad \Pr\left[error|\{\mathbf{z}_0, \mathbf{z}_1\}\right]. \tag{4.3}$$

We now proceed with the analysis of the worst-case jammer function in Gaussian setup.

4.1. Optimal Jammer Behavior Conditioned On The Partial Information

In this part, the optimal behavior in terms of the jammer, i.e. the worst-case jammer behavior defined by (4.3), is derived for the Gaussian setup with the statistical

characterization given in Sec. 2.2. Recall that we assume the jammer function is a potentially random function of $i, \mathbf{x}_0, \mathbf{x}_1$ and $\mathbf{T}$, i.e. $\mathbf{u} = J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$ does not necessarily have a degenerate probability distribution. Theorem 4.1.1 gives us the construction of this optimal jammer function.

Theorem 4.1.1 *Worst-case jammer function criterion for communication with partial information in Gaussian setup with power constraint P given in (4.3) is satisfied by the* deterministic *function*

$$\mathbf{u}^* = \begin{cases} -\sqrt{P}\frac{\mathbf{T}^T(\mathbf{z}_i - \mathbf{z}_j)}{\|\mathbf{T}^T(\mathbf{z}_i - \mathbf{z}_j)\|} & \text{if } \mathbf{z}_i \neq \mathbf{z}_j \\ \mathbf{0} & \text{if } \mathbf{z}_i = \mathbf{z}_j \end{cases} \tag{4.4}$$

where $j, i \in \{0, 1\}, j \neq i$. Moreover, $\mathbf{u}$ satisfies (4.3) if and only if it satisfies (4.4) almost surely for $\mathbf{z}_0 \neq \mathbf{z}_1$ and $\|\mathbf{u}\|^2 \leq P$ for $\mathbf{z}_0 = \mathbf{z}_1$. Also, the worst-case conditional probability of error is given by

$$\Pr\left[error|\{\mathbf{z}_0, \mathbf{z}_1\}\right]\Big|_{\mathbf{u}=\mathbf{u}^*} = \begin{cases} Q\left(\frac{-\sqrt{P}\|\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\| + \|\mathbf{z}_0 - \mathbf{z}_1\|^2/2}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\|}\right) & \text{if } \mathbf{z}_0 \neq \mathbf{z}_1 \\ \frac{1}{2} & \text{if } \mathbf{z}_0 = \mathbf{z}_1 \end{cases} \tag{4.5}$$

Proof:

See Appendix G. ∎

The above theorem establishes the worst-case jammer behavior in Gaussian setup *for each realization of transmission*. Now we proceed to characterize optimal linear transforms under the worst-case jammer behavior *in terms of the statistical structure of the system*.

4.2. Worst-Case Linear Operators In The Probabilistic Sense

In this part, we study the characteristics and design of linear transformations that improve the worst-case failure probability of the binary minimum Euclidean distance detector. We first analyze the worst-case failure probability, defined by (4.2), in terms of its analytical tractability for designing optimal transforms that minimize it. Then we introduce a criterion relating the improved worst-case failure probability to the design of linear transforms constructed, with respect to the statistical structure imposed by the system, in order extract partial information from the transmitter codebook.

In the following remark we study the tractability of P_{α,J^*}.

Remark 4.2.1 *Consider the definition of P_{α,J^*} given by (4.2). If we substitute the second result of Theorem 4.1.1 given by (4.5), we get*

$$\begin{aligned} P_{\alpha,J^*} &= \Pr\left[Q\left(s\left(\mathbf{z}_0,\mathbf{z}_1\right)\right)>\alpha \,\middle|\, \mathbf{z}_0\neq\mathbf{z}_1\right]\Pr[\mathbf{z}_0\neq\mathbf{z}_1]+\Pr\left[\frac{1}{2}>\alpha\right]\Pr[\mathbf{z}_0=\mathbf{z}_1] \\ &= \Pr\left[s\left(\mathbf{z}_0,\mathbf{z}_1\right)<Q^{-1}\left(\alpha\right) \,\middle|\, \mathbf{z}_0\neq\mathbf{z}_1\right] \qquad (4.6) \\ &= \Pr_{\gamma}\left[\frac{-\sqrt{P}\|\mathbf{T}^T\gamma\|+\|\gamma\|^2/2}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|}<Q^{-1}\left(\alpha\right) \,\middle|\, \gamma\neq 0\right] \qquad (4.7) \end{aligned}$$

where $s\left(\mathbf{z}_0,\mathbf{z}_1\right)=\frac{-\sqrt{P}\|\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)\|+\|\mathbf{z}_0-\mathbf{z}_1\|^2/2}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)\|}$ and $\gamma=\mathbf{z}_0-\mathbf{z}_1$. Here, (4.6) follows from the fact that $\Pr[\mathbf{z}_0=\mathbf{z}_1]=0$ and (4.7) is just a restatement. This yields that in order to analyze the above quantity, we need to find the cumulative distribution function of the random variable $\frac{-\sqrt{P}\|\mathbf{T}^T\gamma\|+\|\gamma\|^2/2}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|}$ where $\gamma\sim\mathcal{N}\left(\mathbf{0},2\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)$. However, if we recall (3.35) in Remark 3.3.1, we can clearly say that the above given quantity is impossible to tract in terms of the design of the linear transform to be constructed to minimize P_{α,J^*}.

Remark 4.2.1 pushes us to set upper bounds on P_{α,J^*} and analyze these instead of the original cost function P_{α,J^*}. Proposition 4.2.1 gives us such a revised cost function.

Proposition 4.2.1 *A bound, which is analytically tractable in terms of* $\mathbf{T}$*, is given by*

$$P_{\alpha,J^*} \leq \mathcal{P}\left(\frac{m}{2}, \frac{\beta^2 \lambda_{max}\left(\mathbf{T}\boldsymbol{\Sigma}_e\mathbf{T}^T\right)}{4\lambda_{min}\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)}\right), \tag{4.8}$$

where $\mathcal{P}(.,.)$ *is the* regularized gamma function *and* $\beta = 2\left[Q^{-1}(\alpha) + \sqrt{\frac{P}{\lambda_{min}(\boldsymbol{\Sigma}_e)}}\right]$.

Proof:

See Appendix H. ∎

Proposition 4.2.1 allows us to introduce a *design criterion* for $\mathbf{T}$ which aims to minimize the bound on worst-case failure probability P_{α,J^*}. This criterion, which uses the strictly increasing character of regularized gamma function in its second argument, is given below as the last result of the thesis.

Definition 4.2.1 *The design criterion for the optimal* $\mathbf{T}$ *minimizing the bound on* P_{α,J^*} *in (4.8) is given by*

$$\mathbf{T}^* \triangleq \underset{\substack{\mathbf{T}\in\mathbb{R}^{m\times n}\\ r(\mathbf{T})=m}}{\operatorname{argmin}} \quad \frac{\lambda_{max}\left(\mathbf{T}\boldsymbol{\Sigma}_e\mathbf{T}^T\right)}{\lambda_{min}\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)} \tag{4.9}$$

We finish the thesis by indicating that this design criterion makes sense since it is set to maximize the "worst-performance" (we mean the smallest eigenvalue) of the information shared between the two sides while minimizing the "best performance" (we mean the largest eigenvalue) of the distortion on it. Here the effect of the jammer is confined to the parameter β that determines the value of optimality by (4.8).

5. CONCLUSIONS

In this thesis, we introduce the concept of communication with partial information. The main idea is that the codebooks used by the transmitter and the receiver are different. This concept is different from the concept of communication with side information, where the utilized codebooks are the same but there is extra information available to one of the communicating parties.

Within the context of communication with partial information, we particularly concentrate on a binary detection theoretic scenario. As such, the receiver acts as a detector, using *dimensionality reduced versions* of the encoder codewords, where the dimensionality reduction is achieved via a linear transform. In the most general case, we take into account the potential presence of a jammer, who introduces a disturbance (jammer signal) to the transmitter output subject to a peak power constraint, prior to an additive colored Gaussian noise channel.

In the first case of interest, we assume that the peak power constraint on the jammer is the zero-triviality, which amounts to the no-jammer case. In that case, we find the optimal (in the sense of probability of error) detection rule and subsequently derive optimal classes of linear transforms in two senses: The first criterion is the expected value of the Chernoff bound on the conditional probability of error of the detector; the second criterion is a certain upper bound on the probability of the conditional error probability (conditioned on the partial information) of the detector's being greater than a constant. In the second case of interest, we assume that there is an active jammer. In that scenario, we first derive the optimal (worst case) jammer signal where the detector is confined to perform minimum-distance detection within the subspace of the partial information. This result accordingly yields a criterion on the choice of the dimensionality-reducing linear transform in the worst case.

Although the scenario of focus here is binary detection, we believe that the proposed "communication with partial information" covers several setups of interest, espe-

cially the cases where there is an inherent asymmetry between the transmitter and the receiver due to the unbalanced limitations on the physical resources, such as memory and computational power. In our future research, we plan to explore various communication theoretic setups where asymmetry is the crucial feature.

APPENDIX A: PROOF OF THEOREM 3.1.1

Throughout the proof, we use the definitions of $\mathbf{y}_i \stackrel{\triangle}{=} \mathbf{x}_i + \mathbf{e}$ for $i \in \{0,1\}$. Accordingly, we use $\mu_{y_i|z_i} = \mathrm{E}\,(y_i|z_i)$ and $\mathbf{\Sigma}_{y_i|z_i} = \mathrm{Cov}\,(y_i|z_i)$. We start with the following lemma.

Lemma A.0.1 *For $i \in \{0,1\}$, conditioned on $\mathbf{z}_i$, $\mathbf{y}_i$ is a normal random vector. Furthermore*

$$\mathbf{\Sigma}_{y|z} = \mathbf{\Sigma}_{y_i|z_i} = \mathbf{\Sigma}_x + \mathbf{\Sigma}_e - \mathbf{\Sigma}_x\mathbf{T}^T(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T)^{-1}\mathbf{T}\mathbf{\Sigma}_x, \tag{A.1}$$

is independent of i and positive definite.

Proof:

The crucial point is to show that, for $i \in \{0,1\}$, $\mathbf{y}_i$ and $\mathbf{z}_i$ are jointly normal with a positive definite covariance matrix. First, consider $\begin{bmatrix} \mathbf{x}_i \\ \mathbf{e} \end{bmatrix} \in \mathbb{R}^{2n}$. Since $\mathbf{x}_i$ and $\mathbf{e}$ are both normal and are independent, they are also jointly normal (see [17]) with zero mean and the covariance matrix of $\mathbf{H} \stackrel{\triangle}{=} \begin{bmatrix} \mathbf{\Sigma}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_e \end{bmatrix} \in \mathbb{R}^{2n\times 2n}$. Note that, $\mathbf{H}$ is clearly positive definite, since for any $\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \in \mathbb{R}^{2n}$ where $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, $\mathbf{v}^T\mathbf{H}\mathbf{v} = \mathbf{v}_1^T\mathbf{\Sigma}_x\mathbf{v}_1 + \mathbf{v}_2^T\mathbf{\Sigma}_e\mathbf{v}_2 \geq 0$ by the positive definiteness of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_e$ (that we assumed). By the same token, $\left[\mathbf{v}_1^T\mathbf{\Sigma}_x\mathbf{v}_1 + \mathbf{v}_2^T\mathbf{\Sigma}_e\mathbf{v}_2 = 0\right] \iff [\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}] \iff [\mathbf{v} = \mathbf{0}]$, yielding the positive definiteness of $\mathbf{H}$.

Now, consider the linear transformation from the normal random vector $\begin{bmatrix} \mathbf{x}_i \\ \mathbf{e} \end{bmatrix} \in \mathbb{R}^{2n}$ to the vector $\begin{bmatrix} \mathbf{y}_i \\ \mathbf{z}_i \end{bmatrix} \in \mathbb{R}^{n+m}$ represented by $\mathbf{F} = \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \\ \mathbf{T} & \mathbf{0}_{m\times n} \end{bmatrix} \in \mathbb{R}^{(n+m)\times 2n}$, where $\mathbf{0}_{m\times n}$ denotes the $m \times n$ zero matrix. This linear transform establishes the normality

of $\begin{bmatrix} \mathbf{y}_i \\ \mathbf{z}_i \end{bmatrix} \in \mathbb{R}^{n+m}$ (by the properties of jointly normal random vectors, see [17]) with zero mean and the covariance matrix of $\mathbf{FHF}^T = \begin{bmatrix} \mathbf{\Sigma}_x + \mathbf{\Sigma}_e & \mathbf{\Sigma}_x T^T \\ \mathbf{T\Sigma}_x & \mathbf{T\Sigma}_x\mathbf{T}^T \end{bmatrix}$. To deduce the positive definiteness of this covariance matrix, i.e., $\mathbf{FHF}^T$, it is sufficient to show that $\mathbf{F}$ is full rank. This stems from the fact that if $\mathbf{F}$ is full rank (i.e., if $r(\mathbf{F}) = m+n$ since $m < n$), for any nonzero vector $\mathbf{s} \in \mathbb{R}^{m+n}$ we have $\mathbf{F}^T\mathbf{s} = \mathbf{w} \neq \mathbf{0} \in \mathbb{R}^{2n}$ since $\mathbf{F}^T$ has a trivial *null-space*, so we end-up with $\mathbf{s}^T\mathbf{FHF}^T\mathbf{s} = \mathbf{w}^T\mathbf{Hw} > 0$ by the positive definiteness of $\mathbf{H}$.

To establish the full-rank property of $\mathbf{F}$ (equivalent to having "$\mathbf{F}^T$ has a trivial null-space"), consider $\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} \in \mathbb{R}^{m+n}$ where $\mathbf{a}_1 \in \mathbb{R}^n$ and $\mathbf{a}_2 \in \mathbb{R}^m$. In this case, $\mathbf{F}^T\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 + \mathbf{T}^T\mathbf{a}_2 \\ \mathbf{a}_1 \end{bmatrix}$. Suppose there exists some $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{F}^T\mathbf{a} = \mathbf{0}$. This implies, $\mathbf{a}_1 = \mathbf{0}$ and $\mathbf{T}^T\mathbf{a}_2 = \mathbf{0}$. However, since $r(\mathbf{T}) = m$, $\left[\mathbf{T}^T\mathbf{a}_2 = \mathbf{0}\right] \iff [\mathbf{a}_2 = \mathbf{0}]$. Therefore, $\left[\mathbf{F}^T\mathbf{a} = \mathbf{0}\right] \iff [\mathbf{a} = \mathbf{0}]$ and hence contradiction. Thus, $\mathbf{F}$ is necessarily full-rank implying positive-definiteness of the covariance matrix of $\begin{bmatrix} \mathbf{y}_i \\ \mathbf{z}_i \end{bmatrix}$, i.e., $\mathbf{FHF}^T$.

The remaining part is easy: normality of $[\mathbf{y}_i \mid \mathbf{z}_i]$ follows from the properties of normal distributed random variables (see [17]). The positive definiteness of the corresponding covariance matrix $\mathbf{\Sigma}_{y_i|z_i} = \mathbf{\Sigma}_x + \mathbf{\Sigma}_e - \mathbf{\Sigma}_x\mathbf{T}^T(\mathbf{T\Sigma}_x\mathbf{T}^T)^{-1}\mathbf{T\Sigma}_x$ follows from the fact that it is the inverse of a principal submatrix of the inverse of $\mathbf{FHF}^T$, which is positive definite (see (7.1.2) and (7.7.5) in [15]). Also, $\mathbf{\Sigma}_{y_i|z_i}$ is clearly independent of $i \in \{0, 1\}$. ∎

Corollary A.0.1 *Per Lemma A.0.1, since $\mathbf{\Sigma}_{y|z}$ is positive definite, it is also invertible. Further, it has an invertible square root.*

Remark A.0.2 *From properties of normal random vectors, we have*

$$\mu_{y_i|z_i} = E(\mathbf{y}_i \,|\, \mathbf{z}_i) = \mathbf{\Sigma}_x \mathbf{T}^T (\mathbf{T}\mathbf{\Sigma}_x \mathbf{T}^T)^{-1} \mathbf{z}_i. \tag{A.2}$$

Using Lemma A.0.1 and Remark A.0.2, given $\mathbf{y}$ is observed we have

$$p\left(\mathbf{y}_i|\mathbf{z}_i\right)\Big|_{y_i=y} = \frac{1}{(2\pi)^{n/2}\det(\mathbf{\Sigma}_{y|z})^{1/2}} \exp\left[-\frac{(\mathbf{y}-\mu_{y_i|z_i})^T \mathbf{\Sigma}_{y|z}^{-1}(\mathbf{y}-\mu_{y_i|z_i})}{2}\right], \tag{A.3}$$

where $\mathbf{\Sigma}_{y|z}$ and $\mu_{y_i|z_i}$ are given in (A.1) and (A.2), respectively. Then, using the above distribution of $[\mathbf{y}_i \,|\, \mathbf{z}_i]$ and that $\mathbf{\Sigma}_{y|z}$ is constant in i (and also that $\det\left(\mathbf{\Sigma}_{y|z}\right) \neq 0$) the maximum likelihood detection rule (3.8) can be written as

$$\exp\left[-\frac{(\mathbf{y}-\mu_{y_0|z_0})^T \mathbf{\Sigma}_{y|z}^{-1}(\mathbf{y}-\mu_{y_0|z_0})}{2}\right] \underset{H_1}{\overset{H_0}{\gtrless}} \exp\left[-\frac{(\mathbf{y}-\mu_{y_1|z_1})^T \mathbf{\Sigma}_{y|z}^{-1}(\mathbf{y}-\mu_{y_1|z_1})}{2}\right] \tag{A.4}$$

which is equivalent to stating

$$\|\mathbf{\Sigma}_{y|z}^{-1/2}\left(\mathbf{y}-\mu_{y_0|z_0}\right)\| \underset{H_0}{\overset{H_1}{\gtrless}} \|\mathbf{\Sigma}_{y|z}^{-1/2}\left(\mathbf{y}-\mu_{y_1|z_1}\right)\| \tag{A.5}$$

since exp(.) is a strictly increasing function of its argument. So, the maximum likelihood detection rule for $P = 0$ is given by (A.5), which is the first result of Theorem 3.1.1.

Now, if we suppose that H_0 is the true hypothesis, by (A.5) we have

$$\Pr\left[\text{error} \,\middle|\, H_0\right] = \Pr\left[\frac{\|\mathbf{\Sigma}_{y|z}^{-1/2}\left(\mathbf{y}-\mu_{y_0|z_0}\right)\|^2}{\|\mathbf{\Sigma}_{y|z}^{-1/2}\left(\mathbf{y}-\mu_{y_1|z_1}\right)\|^2} > 1 \;\middle|\; \mathbf{y} \sim \mathcal{N}\left(\mu_{y_0|z_0}, \mathbf{\Sigma}_{y|z}\right)\right] \tag{A.6}$$

After some straightforward algebra, from (A.6) we get

$$\Pr\left[\text{error} \,\middle|\, H_0\right] = \Pr\left[\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}\mathbf{y} < \frac{c}{2} \;\middle|\; \mathbf{y} \sim \mathcal{N}(\mu_{y_0|z_0}, \boldsymbol{\Sigma}_{y|z})\right], \tag{A.7}$$

where we let $c = \|\boldsymbol{\Sigma}_{y|z}^{-1/2}\mu_{y_0|z_0}\|^2 - \|\boldsymbol{\Sigma}_{y|z}^{-1/2}\mu_{y_1|z_1}\|^2$. Here, the random variable $\theta \in \mathbb{R}$ defined by $\theta \triangleq \left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}\mathbf{y}$ is normally-distributed (since $\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$) with mean and variance (conditioned on H_0) given respectively by

$$\mu_\theta = \left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}\mu_{y_0|z_0}, \tag{A.8}$$

$$\sigma_\theta^2 = \left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right). \tag{A.9}$$

Then, $\Pr\left[\text{error} \,|\, H_0\right]$ is given in terms of the standard Q-function. As a result, we have

$$\Pr\left[\text{error} \,|\, H_0\right] = \mathrm{Q}\left(\frac{\left[\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)^T \boldsymbol{\Sigma}_{y|z}^{-1}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)\right]^{\frac{1}{2}}}{2}\right), \tag{A.10}$$

$$= \mathrm{Q}\left(\frac{\|\boldsymbol{\Sigma}_{y|z}^{-1/2}\left(\mu_{y_0|z_0} - \mu_{y_1|z_1}\right)\|}{2}\right). \tag{A.11}$$

Furthermore, from symmetry with respect to hypotheses and equal priors, we have $\Pr\left[\text{error} \,|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right] = \Pr\left[\text{error} \,|\, H_0\right]$; hence the proof. □

APPENDIX B: PROOF OF PROPOSITION 3.2.1

We first show the first result, i.e., the Chernoff bound on $\Pr[\text{error} \mid \{\mathbf{z}_0, \mathbf{z}_1\}]$. The crucial point is to see that the argument of the Q-function in (3.10) is nonnegative (by the positive definiteness of $\mathbf{\Sigma}_{y \mid z}$). The result is, then, obvious by stating the Chernoff bound in [16], i.e., $\mathrm{Q}(x) \leq \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$ for $x \geq 0$ and second result of Theorem 3.1.1.

To show the second part, we first restate the first result in a more suitable way:

$$\Pr[\text{error} \mid \{\mathbf{z}_0, \mathbf{z}_1\}] \leq \frac{1}{2}\exp\left(-\frac{[\mathbf{A}\gamma]^T \mathbf{B}^{-1}[\mathbf{A}\gamma]}{8}\right) \tag{B.1}$$

where we use the result of Remark A.0.2 and define $\mathbf{A} \stackrel{\triangle}{=} \mathbf{\Sigma}_x \mathbf{T}^T (\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T)^{-1}$, $\mathbf{B} \stackrel{\triangle}{=} \mathbf{\Sigma}_{y \mid z}$ and $\gamma \stackrel{\triangle}{=} \mathbf{z}_0 - \mathbf{z}_1$. This restatement allows us to state the Chernoff bound as a function of the random vector $\gamma \in \mathbb{R}^m$, hence the $m \times m$-fold integration (integration on $\mathbb{R}^{m \times m}$) to evaluate the expected bound reduces to an m-fold integration (integration on $\mathbb{R}^m$). Here, it is clear that $\gamma \sim \mathcal{N}(\mathbf{0}, 2\mathbf{\Sigma}_z)$, where, by properties of normal random vectors, $\mathbf{\Sigma}_z = \mathrm{Cov}(\mathbf{z}_0) = \mathrm{Cov}(\mathbf{z}_1) = \mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T$ (see [17]).

Now, let $\mathrm{g}(\mathbf{T}, \mathbf{z}_0, \mathbf{z}_1) \stackrel{\triangle}{=} \frac{1}{2}\exp\left(-\frac{[\mathbf{A}\gamma]^T\mathbf{B}^{-1}[\mathbf{A}\gamma]}{8}\right)$. Then, by the above notation, we get $\mathrm{g}(\mathbf{T}, \mathbf{z}_0, \mathbf{z}_1) = \mathrm{g}(\mathbf{T}, \gamma)$, yielding

$$\begin{aligned} \mathrm{E}_{\{\mathbf{z}_0, \mathbf{z}_1\}}(\mathrm{g}(\mathbf{T}, \mathbf{z}_0, \mathbf{z}_1)) &= \mathrm{E}_\gamma(\mathrm{g}(\mathbf{T}, \gamma)) && \text{(B.2)} \\ &= \int_{\mathbb{R}^m} \mathrm{g}(\mathbf{T}, \gamma) p_\gamma(\gamma)\, \mathrm{d}\gamma\ , && \text{(B.3)} \end{aligned}$$

whenever it exists ($p_\gamma(\gamma)$ denotes the probability density function of γ). To show the existence of this expectation, and its equivalence to the second result of Proposi-

tion 3.2.1, we simply evaluate it. By the above statement, we have

$$\mathrm{E}_{\gamma}\left(\mathrm{g}(\mathbf{T},\gamma)\right) = \int_{\mathbb{R}^m} \frac{1/2}{(2\pi)^{\frac{m}{2}}\det\left(2\mathbf{\Sigma}_z\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\gamma^T\hat{\mathbf{\Sigma}}_{\gamma}\gamma\right)\mathrm{d}\gamma \tag{B.4}$$

$$= \frac{1}{2}\left(\frac{\det\left(\left[\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}\right]^{-1}\right)}{\det\left(2\mathbf{\Sigma}_z\right)}\right)^{\frac{1}{2}} \tag{B.5}$$

$$= \frac{1}{2}\left\{\det\left(\mathbf{I}_m+\frac{\mathbf{\Sigma}_z\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{2}\right)\right\}^{-\frac{1}{2}}, \tag{B.6}$$

where $\hat{\mathbf{\Sigma}}_{\gamma} = \left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}$ and (B.5) result from the fact that the remaining term is the integration of the probability distribution function of a normal random vector with zero mean and the covariance matrix of $\left[\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}\right]^{-1}$ on $\mathbb{R}^m$. The convergence of this integral is guaranteed by the positive definiteness of the covariance matrix $\left[\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}\right]^{-1}$, which can be shown in the following way: First, by assumption we have that $\mathbf{\Sigma}_x$ is positive definite and $\mathbf{T}$ is full-rank. This yields (see the proof of the positive definiteness of $\mathbf{FHF}^T$ in the proof of Lemma A.0.1) $\left[\text{the positive definiteness of } \mathbf{\Sigma}_z=\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right] \iff \left[\text{the positive definiteness of } 2\mathbf{\Sigma}_z\right] \iff \left[\text{the positive definiteness of } \left(2\mathbf{\Sigma}_z\right)^{-1}\right]$ (see [15]). Similarly, since by Lemma A.0.1 $\mathbf{B}$ is positive definite (and it is easy to show that, $\mathbf{A}\in\mathbb{R}^{n\times m}$ is full rank, so we skip this here), we have $[\mathbf{B}^{-1}$ is positive definite$] \implies \left[\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A} \text{ is positive definite}\right] \iff \left[\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}/4 \text{ is positive definite}\right]$. In short, we get $\left[\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}\right]^{-1}$ is positive definite since it is the inverse of the sum of two positive definite matrices, which is itself positive definite. This result allows to make sense of the inverse of this covariance matrix, i.e., $\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}$ and conclude that the integral of the probability density function of the normal random vector with covariance $\left[\left(2\mathbf{\Sigma}_z\right)^{-1}+\frac{\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}}{4}\right]^{-1}$ converges to 1. Also (B.6) follows from the properties of determinant (see [15]). Hence the second result of theorem follows by substituting $\mathbf{\Sigma}_z$, $\mathbf{A}$ and $\mathbf{B}$ using their definitions.

□

APPENDIX C: PROOF OF PROPOSITION 3.2.2

We first define $\mathcal{J}(\mathbf{T}) \triangleq \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\right)$ to denote the objective function in (3.17). Then by Theorem 3.1.1, we have

$$\mathcal{J}(\mathbf{T}) = \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{T}\boldsymbol{\Sigma}_x\left(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\mathbf{T}\boldsymbol{\Sigma}_x\right)^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\right), \tag{C.1}$$

where we let $\boldsymbol{\Sigma} \triangleq \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_e$.

Now, since $\boldsymbol{\Sigma}_x$ is symmetric, by spectral decomposition of $\boldsymbol{\Sigma}_x$ (see [15]) we let $\boldsymbol{\Sigma}_x = \mathbf{F}\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_x}\mathbf{F}^T = \mathbf{F}\boldsymbol{\Lambda}^2\mathbf{F}^T$, where $\mathbf{F}$ is unitary and $\boldsymbol{\Lambda}$ is diagonal with *positive* entries. Here, the second equality follows from the fact that $\boldsymbol{\Lambda}_{\boldsymbol{\Sigma}_x}$ is diagonal with positive entries, since $\boldsymbol{\Sigma}_x$ is positive definite by assumption, and hence it is also positive definite with a diagonal square root with positive entries, which we denote here by $\boldsymbol{\Lambda}$. Now, if we let $\mathbf{K} \triangleq \mathbf{T}\mathbf{F}\boldsymbol{\Lambda}$, we can write $\mathcal{J}(\mathbf{T})$ as

$$\mathcal{J}(\mathbf{T}) = \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{K}\boldsymbol{\Lambda}\mathbf{F}^T\left(\boldsymbol{\Sigma} - \mathbf{F}\boldsymbol{\Lambda}\left(\mathbf{K}^T\left(\mathbf{K}\mathbf{K}^T\right)^{-1}\mathbf{K}\right)\boldsymbol{\Lambda}\mathbf{F}^T\right)^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{K}^T\left(\mathbf{K}\mathbf{K}^T\right)^{-1}\right). \tag{C.2}$$

Then, for $\mathbf{K} \in \mathbb{R}^{m\times n}$ if we use SVD given by (2.1) we have $\mathbf{K} = \mathbf{U}_{\mathbf{K}}\boldsymbol{\Lambda}_{\mathbf{K}}\mathbf{V}_{\mathbf{K}}^T$, where $\mathbf{U}_{\mathbf{K}} \in \mathbb{R}^{m\times m}$ is unitary and $\mathbf{V}_{\mathbf{K}} \in \mathbb{R}^{m\times n}$ is orthonormal, since $r(\mathbf{K}) = n - dim\left(\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{K}\mathbf{x} = \mathbf{0}\}\right) = n - dim\left(\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{T}\mathbf{x} = \mathbf{0}\}\right) = r(\mathbf{T}) = m$. So we get, $\mathbf{K}^T\left(\mathbf{K}\mathbf{K}^T\right)^{-1}\mathbf{K} = \mathbf{V}_{\mathbf{K}}\mathbf{V}_{\mathbf{K}}^T$ and $\mathbf{K}^T\left(\mathbf{K}\mathbf{K}^T\right)^{-1} = \mathbf{V}_{\mathbf{K}}\boldsymbol{\Lambda}_{\mathbf{K}}^{-1}\mathbf{U}_{\mathbf{K}}^T$. Then $\mathcal{J}(\mathbf{T})$ becomes

$$\mathcal{J}(\mathbf{T}) = \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{U}_{\mathbf{K}}\boldsymbol{\Lambda}_{\mathbf{K}}\mathbf{V}_{\mathbf{K}}^T\left(\boldsymbol{\Lambda}^{-1}\mathbf{F}^T\boldsymbol{\Sigma}\mathbf{F}\boldsymbol{\Lambda}^{-1} - \mathbf{V}_{\mathbf{K}}\mathbf{V}_{\mathbf{K}}^T\right)^{-1}\mathbf{V}_{\mathbf{K}}\boldsymbol{\Lambda}_{\mathbf{K}}^{-1}\mathbf{U}_{\mathbf{K}}^T\right), \tag{C.3}$$

where we make use of invertibility of $\mathbf{F}, \mathbf{\Lambda}$ and $\mathbf{F}^T$. Let $\mathbf{P} \stackrel{\triangle}{=} \mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}\mathbf{F}\mathbf{\Lambda}^{-1}$, then taking into parentheses from both sides we get

$$\begin{aligned} \mathcal{J}(\mathbf{T}) &= \det(\mathbf{U_K}\mathbf{\Lambda_K}) \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{V_K}^T\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right)^{-1}\mathbf{V_K}\right) \det\left(\mathbf{\Lambda_K}^{-1}\mathbf{U_K}^T\right) \quad \text{(C.4)} \\ &= \det\left(\mathbf{I}_m + \frac{1}{2}\mathbf{V_K}^T\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right)^{-1}\mathbf{V_K}\right), \quad \text{(C.5)} \end{aligned}$$

by the properties of det(.) and that $\det(\mathbf{U_K}\mathbf{\Lambda_K}) = \left[\det\left(\mathbf{\Lambda_K}^{-1}\mathbf{U_K}^T\right)\right]^{-1}$. Moreover, using *matrix inversion lemma* (see [15]) yields

$$\begin{aligned} \mathbf{I}_m + \mathbf{V_K}^T\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right)^{-1}\mathbf{V_K} &= \left(\mathbf{I}_m - \mathbf{V_K}^T\mathbf{P}^{-1}\mathbf{V_K}\right)^{-1} \quad \text{(C.6)} \\ \Longrightarrow \mathcal{J}(\mathbf{T}) &= \left(\frac{1}{2}\right)^m \det\left(\mathbf{I}_m + \left(\mathbf{I}_m - \mathbf{V_K}^T\mathbf{P}^{-1}\mathbf{V_K}\right)^{-1}\right). \quad \text{(C.7)} \end{aligned}$$

Remark C.0.3 *For (C.6) to hold, both of the inverses must necessarily exist. Here,* $\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right)^{-1}$ *exists because* $\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right) > 0$. *To see this, we write* $\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T = \left(\mathbf{I}_n - \mathbf{V_K}\mathbf{V_K}^T\right) + \mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}_e\mathbf{F}\mathbf{\Lambda}^{-1}$ *by the definition of* $\mathbf{P}$. *Then, since* $\mathbf{\Sigma}_e > 0$ *and* $\mathbf{\Lambda}^{-1}\mathbf{F}^T$ *is invertible* $\mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}_e\left(\mathbf{\Lambda}^{-1}\mathbf{F}^T\right)^T > 0$, *clearly. Also,* $\mathbf{I}_n - \mathbf{V_K}\mathbf{V_K}^T \geq 0$ *since* $\lambda_i\left(\mathbf{I}_n - \mathbf{V_K}\mathbf{V_K}^T\right) = 1 - \lambda_i\left(\mathbf{V_K}\mathbf{V_K}^T\right)$ *for all* $i \in \{1, 2, ..., n\}$. *This is because* $\mathbf{V_K}\mathbf{V_K}^T\mathbf{x} = \lambda\mathbf{x}$ *implies (by multiplying both sides with* $\mathbf{V_K}^T$*)* $\mathbf{V_K}^T\mathbf{x} = \lambda\mathbf{V_K}^T\mathbf{x}$ *yielding* $\lambda = 1$ *or* $\lambda = 0$. *So* $\left(\mathbf{P} - \mathbf{V_K}\mathbf{V_K}^T\right)$, *which is the sum of a positive definite and a nonnegative definite matrix , is positive definite. Moreover,* $\left(\mathbf{I}_m - \mathbf{V_K}^T\mathbf{P}^{-1}\mathbf{V_K}\right)^{-1}$ *exists because* $\mathbf{I}_m - \mathbf{V_K}^T\mathbf{P}\mathbf{V_K} > 0$. *To deduce this fact, consider* $\mathbf{I}_m - \mathbf{V_K}^T\mathbf{P}^{-1}\mathbf{V_K} = \mathbf{V_K}^T\left(\mathbf{I}_n - \mathbf{P}^{-1}\right)\mathbf{V_K}$ *since* $\mathbf{V_K}$ *is orthonormal. Here,* $\mathbf{I}_n - \mathbf{P}^{-1}$ *is positive definite because, for all* $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$,

$$\begin{aligned} \mathbf{x}^T\left(\mathbf{I}_n - \mathbf{P}^{-1}\right)\mathbf{x} &\geq \|\mathbf{x}\|^2\left(1 - \lambda_{max}\left(\mathbf{P}^{-1}\right)\right) \quad \text{(C.8)} \\ &= \|\mathbf{x}\|^2\left(1 - 1/\lambda_{min}\left(\mathbf{P}\right)\right) \quad \text{(C.9)} \\ &= \|\mathbf{x}\|^2\left(1 - 1/\lambda_{min}\left(\mathbf{I}_n + \mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}_e\mathbf{F}\mathbf{\Lambda}^{-1}\right)\right) \quad \text{(C.10)} \\ &= \|\mathbf{x}\|^2\left(1 - 1/\left(1 + \lambda_{min}\left(\mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}_e\mathbf{F}\mathbf{\Lambda}^{-1}\right)\right)\right) \quad \text{(C.11)} \\ &> 0 \quad \text{(C.12)} \end{aligned}$$

since $\mathbf{\Lambda}^{-1}\mathbf{F}^T\mathbf{\Sigma}_e\mathbf{F}\mathbf{\Lambda}^{-1} > 0$. *So,* $\mathbf{I}_m - \mathbf{V}_\mathbf{K}^T\mathbf{P}^{-1}\mathbf{V}_\mathbf{K} = \mathbf{V}_\mathbf{K}^T\left(\mathbf{I}_n - \mathbf{P}^{-1}\right)\mathbf{V}_\mathbf{K} > 0$ *since* $r\left(\mathbf{V}_\mathbf{K}\right) = r\left(\mathbf{K}\right) = m$, *i.e.,* $\mathbf{V}_\mathbf{K}$ *is full rank.*

Remark C.0.4 *(C.7) gives the first reduction in the problem of constructing optimal transform* $\mathbf{T}^*$, *i.e., the original problem reduces to characterizing the orthonormal* $\mathbf{V}_\mathbf{K}$ *that maximizes* $\mathcal{J}\left(\mathbf{T}\right)$ *in (C.7).*

Now, if we let $\mathbf{P} = \mathbf{U}_\mathbf{P}\mathbf{\Lambda}_\mathbf{P}\mathbf{U}_\mathbf{P}^T$, by spectral decomposition, $\hat{\mathbf{\Lambda}}_\mathbf{P} \triangleq \mathbf{I}_n - \mathbf{\Lambda}_\mathbf{P}$ and $\mathbf{M} \triangleq \mathbf{U}_\mathbf{P}^T\mathbf{V}_\mathbf{K}$, which is orthonormal since $\mathbf{U}_\mathbf{P}$ is unitary and $\mathbf{V}_\mathbf{K}$ is orthonormal, we can rewrite $\mathcal{J}\left(\mathbf{T}\right)$ as

$$\mathcal{J}\left(\mathbf{T}\right) = \left(\frac{1}{2}\right)^m \det\left(\mathbf{I}_m + \left(\mathbf{M}^T\left(\mathbf{I}_n - \mathbf{\Lambda}_\mathbf{P}^{-1}\right)\mathbf{M}\right)^{-1}\right) \tag{C.13}$$

$$= \left(\frac{1}{2}\right)^m \prod_{i=1}^{m} \quad 1 + \frac{1}{\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)} \tag{C.14}$$

Then, if we suppose that there exists $\mathbf{M}^* \in \mathcal{S}_\mathbf{M}$ maximizing $\mathcal{J}\left(\mathbf{T}\right)$ in (C.14) then we can find a $\mathbf{T} \in \mathcal{S}_\mathbf{T}$ given by $\mathbf{T} = \mathbf{U}_\mathbf{K}\mathbf{\Lambda}_\mathbf{K}\mathbf{M}^{*^T}\mathbf{U}_\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{F}^T$ (by the definitions of $\mathbf{M}$ and $\mathbf{K}$ and by the SVD of $\mathbf{K}$ and $\mathbf{P}$). Now, is this $\mathbf{T}$ optimal in the sense of Definition 3.2.2? We show that $\mathbf{M}^*$ corresponds to $\mathbf{T}^*$ by the following Lemma.

Lemma C.0.2 *Let* $\mathbf{M}^*$ *be given by (3.18). Then* $\mathbf{T} = \mathbf{U}_\mathbf{K}\mathbf{\Lambda}_\mathbf{K}\mathbf{M}^{*^T}\mathbf{U}_\mathbf{P}^T\mathbf{\Lambda}^{-1}\mathbf{F}^T$ *satisfies (3.17), i.e.,* $\mathbf{T} = \mathbf{T}^*$.

Proof:

First of all, let $\mathcal{G}\left(\mathbf{M}\right) \triangleq \left(\frac{1}{2}\right)^m \prod_{i=1}^{m} 1 + \frac{1}{\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)}$. Then by (C.14), $\mathcal{G}\left(\mathbf{M}\right) = \mathcal{J}\left(\mathbf{T}\right)$, where we find $\mathbf{M}$ from $\mathbf{T}$ by first taking the SVD of $\mathbf{K} = \mathbf{T}\mathbf{F}\mathbf{\Lambda}$ and then multiplying the matrix of right singular vectors of $\mathbf{K}$ by $\mathbf{U}_\mathbf{P}$, which is given. Then we define the mapping that corresponds to the relation between $\mathbf{M}$ and $\mathbf{T}$. Let $f : \mathcal{S}_\mathbf{T} \rightarrow \mathcal{S}_\mathbf{M}$, $f\left(\mathbf{T}\right) = \mathbf{M} = \mathbf{U}_\mathbf{P}^T\mathbf{\Lambda}\mathbf{F}^T\mathbf{T}^T\mathbf{U}_\mathbf{K}\mathbf{\Lambda}_\mathbf{K}^{-1}$. This definition is equivalent to $\mathbf{T} = \mathbf{U}_\mathbf{K}\mathbf{\Lambda}_\mathbf{K}\mathbf{M}^{*^T}\mathbf{U}_\mathbf{P}^T\mathbf{\Lambda}^{-1}\mathbf{F}^T$, by

the invertibility of $\mathbf{U_K}, \mathbf{\Lambda_K}, \mathbf{U_P}, \mathbf{\Lambda}^{-1}$ and $\mathbf{F}^T$ (if $\mathbf{U_K}, \mathbf{\Lambda_K}$ are given). This mapping has several properties. One of these properties is that for any $\mathbf{T} \in \mathcal{S_T}$ there exists unique $\mathbf{M} = f(\mathbf{T}) \in \mathcal{S_M}$, since $\mathbf{U_P}, \mathbf{F}$ and $\mathbf{\Lambda}$ are invertible (and given) and SVD of $\mathbf{K}$ is unique (see Remark C.0.5). Also, for each $\mathbf{M} \in \mathcal{S_M}$ there exists $\mathbf{T} \in f^{-1}(\mathbf{M})$, i.e. in the pre-image of f, given by $\mathbf{T} = \mathbf{U_K}\mathbf{\Lambda_K}\mathbf{M}^{*^T}\mathbf{U_P}^T\mathbf{\Lambda}^{-1}\mathbf{F}^T$, however for given $\mathbf{U_P}, \mathbf{F}$ and $\mathbf{\Lambda}$ and $\mathbf{M}$, $\mathbf{T} \in f^{-1}(\mathbf{M})$ is not unique, since we can choose any unitary $\mathbf{U_K} \in \mathbb{R}^{m \times m}$ and any non-singular diagonal $\mathbf{\Lambda_K} \in \mathbb{R}^{m \times m}$ to satisfy $f(\mathbf{T}) = \mathbf{M}$. Therefore, we have characterized f as an *onto function* from $\mathcal{S_T}$ to $\mathcal{S_M}$.

Now suppose $\mathbf{T}^*$, given by (3.17), satisfies $\mathbf{T}^* \notin f^{-1}(\mathbf{M}^*)$, where $\mathbf{M}^*$ is given by (3.18). Here, since f is onto, $f^{-1}(\mathbf{M}^*) \neq \emptyset$. Also, since f is a function we have $\mathcal{J}(\mathbf{T}^*) = \mathcal{G}(f(\mathbf{T}^*)) \triangleq \mathcal{G}(\mathbf{M}^{**})$, where, given $\mathbf{T}^*$, $\mathbf{M}^{**}$ is a *unique* value satisfying $\mathbf{M}^{**} = \mathbf{U_P}^T\mathbf{\Lambda}\mathbf{F}^T\mathbf{T}^{*^T}\mathbf{U_K}\mathbf{\Lambda_K}^{-1} \neq \mathbf{M}^*$, by assumption. Also, again since f is a function there exists no $\mathbf{T} \in f^{-1}(\mathbf{M}^*)$ satisfying $f(\mathbf{T}) = \mathbf{M}^{**}$. Consequently we have

$$\left[\mathcal{J}(\mathbf{T}^*) > \mathcal{J}(\mathbf{T}) \quad \forall \mathbf{T} \in f^{-1}(\mathbf{M}^*)\right] \quad \Longrightarrow \quad \left[\mathcal{G}(\mathbf{M}^{**}) > \mathcal{G}(\mathbf{M}^*)\right], \tag{C.15}$$

where the left side of (C.15) follows from definition of $\mathbf{T}^*$. So we are done by contradiction and observing

$$f^{-1}(\mathbf{M}^*) = \{\mathbf{T} \in \mathcal{S_T} | \mathbf{T} = \mathbf{U_K}\mathbf{\Lambda_K}\mathbf{M}^{*^T}\mathbf{U_P}^T\mathbf{\Lambda}^{-1}\mathbf{F}^T \text{ for unitary } \mathbf{U_K}, \text{ diagonal } \mathbf{\Lambda_K}\}. \tag{C.16}$$

∎

Remark C.0.5 *In the above proof of Lemma C.0.2, if we do not assume* exact *uniqueness but uniqueness up-to ordering for SVD, proof still remains valid. This is because of the fact that this only results in the failure for f to have a unique value for given $\mathbf{T}$. However, this may result in only the existence of some $\mathbf{T} \in f^{-1}(\mathbf{M}^*)$ satisfying $\{\mathbf{M}^*, \mathbf{M}^{**}\} \in f(\mathbf{T})$, may not result in $\mathbf{M}^* \in f(\mathbf{T}^*)$ (since $\mathbf{T}^* \notin f^{-1}(\mathbf{M}^*)$). But, if $\mathbf{M}^{**} \in f(\mathbf{T})$ for some $\mathbf{T} \in f^{-1}(\mathbf{M}^*)$, then $\mathbf{M}^*$ and $\mathbf{M}^{**}$ are permutated versions of each other, hence $\mathbf{U_K}$ used to construct $\mathbf{T}^*$ from $\mathbf{M}^{**}$ can be permutated to yield*

another unitary matrix $\hat{\mathbf{U}}_{\mathbf{K}}$ *to construct* $\mathbf{T}^*$ *from* $\mathbf{M}^*$*, which yields a contradiction. More important than this is that, if the value for* $f(\mathbf{T})$ *is not unique, it yields a set of permutations of an orthogonal matrix* $\mathbf{M}$ *that we take as the value of* $f(\mathbf{T})$ *for non-decreasing singular values case. However, by Proposition 3.2.3 we show that the cost function* $\mathcal{G}(\mathbf{M})$ *is invariant under permutations, therefore we still have* $\mathcal{J}(\mathbf{T}) = \mathcal{G}(\mathbf{M})$ *without any ambiguity. So we suppose* f *to be an onto function for the simplicity of the proof of Lemma C.0.2.*

Lemma C.0.2 gives the existence of $\mathbf{T}^*$, with the sufficiency of the existence of $\mathbf{M}^*$, and also provides its form depending on $\mathbf{M}^*$. In this form, the matrices $\mathbf{U}_{\mathbf{K}} \in \mathbb{R}^{m\times m}$ and $\mathbf{\Lambda}_{\mathbf{K}} \in \mathbb{R}^{m\times m}$, which are respectively unitary and non-singular diagonal, can be chosen arbitrarily to form $\mathbf{T}^*$ from $\mathbf{M}^*$. So, as a final result, we get the cardinality of the set of optimal transforms $\mathbf{T}$ equal to that of $\mathbb{R}^{m\times m}\times\mathbb{R}^{m\times m}\times\{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}|\mathbf{M}=\mathbf{M}^*\}$. Supposing the existence of $\mathbf{M}^*$, i.e. $\{\mathbf{M}\in\mathcal{S}_{\mathbf{M}}|\mathbf{M}=\mathbf{M}^*\}\neq\emptyset$, this cardinality is at least equal to that of $\mathbb{R}^{m\times m}\times\mathbb{R}^{m\times m}$, and is at most equal to that of $\mathbb{R}^{m\times m}\times\mathbb{R}^{m\times m}\times\mathbb{R}^{m\times n}$, i.e. is equal to that of the continuum (see [18]). □

APPENDIX D: PROOF OF PROPOSITION 3.2.3

First, consider the expression in (3.18). This expression is simply the product of positive real numbers, since $\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right) > 0$ for all $i \in \{1,...,m\}$ by the positive definiteness of $\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}$ (because orthonormal $\mathbf{M}$ is full-rank and $\hat{\mathbf{\Lambda}}_\mathbf{P} > 0$, see Remark C.0.3). So, we follow the strategy of maximizing each positive factor of $\prod_{i=1}^{m} 1 + \frac{1}{\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)}$ over $\mathcal{S}_\mathbf{M}$ in order to maximize the resulting product $\mathcal{G}(\mathbf{M})$ ($\mathcal{S}_\mathbf{M}$ and $\mathcal{G}(\mathbf{M})$ are defined in Proposition 3.2.2 and in Appendix C, respectively). This, clearly, corresponds to minimizing $\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)$ for all $i \in \{1,...,m\}$ over $\mathcal{S}_\mathbf{M}$. Here the following theorem, known as the *Poincaré seperation theorem* that is a result of the *inclusion principle*, is used to construct $\mathbf{M}$ minimizing $\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)$ for all $i \in \{1,...,m\}$.

Theorem D.0.1 *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric, and let m be a given integer with $1 \leq m \leq n$, and $\mathbf{B}_m = \mathbf{U}^T\mathbf{A}\mathbf{U}$, where $\mathbf{U} \in \mathbb{R}^{n\times m}$ is orthonormal. If eigenvalues of $\mathbf{A}$ and $\mathbf{B}_m$ are arranged in non-decreasing order, we have*

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{B}_m) \leq \lambda_{i+n-m}(\mathbf{A}) \quad i = 1,2,...,m \tag{D.1}$$

Proof:

See pp. $190-191$ of [15]. ∎

Then by (D.1) we have $\lambda_i\left(\hat{\mathbf{\Lambda}}_\mathbf{P}\right) \leq \lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)$ for all $i \in \{1,2,...,m\}$ and for all $\mathbf{M} \in \mathcal{S}_\mathbf{M}$, where $\lambda_i\left(\hat{\mathbf{\Lambda}}_\mathbf{P}\right)$ is a given quantity, i.e. constant in $\mathbf{M}$ for all i. Now consider $\mathbf{M} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix} \in \mathcal{S}_\mathbf{M}$. Since $\hat{\mathbf{\Lambda}}_\mathbf{P}$ is diagonal with positive entries in the diagonal, we clearly have $\lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right) = \lambda_i\left(\hat{\mathbf{\Lambda}}_\mathbf{P}\right)$ for all $i \in \{1,2,...,m\}$, i.e. $\mathbf{M} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix} \in \mathcal{S}_\mathbf{M}$ achieves the lower bound in (D.1) and hence we conclude $\mathbf{M} \in \{\mathbf{M} \in \mathcal{S}_\mathbf{M} \mid \mathbf{M} = \mathbf{M}^*\}$. On the other hand, we know that eigenvalues are

invariant to multiplications with tranposes of unitary matrices from left and unitary matrices from right (see [15]), i.e. for any $\mathbf{\Gamma}_m \in \mathbb{R}^{m\times m}$ that is unitary and for any $i \in \{1,2,...,m\}$, $\lambda_i\left(\mathbf{\Gamma}_m^T\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\mathbf{\Gamma}_m\right) = \lambda_i\left(\mathbf{M}^T\hat{\mathbf{\Lambda}}_\mathbf{P}\mathbf{M}\right)$. Then, if $\mathbf{M} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix} \in \mathcal{S}_\mathbf{M}$, which satisfies $\mathbf{M} \in \{\mathbf{M} \in \mathcal{S}_\mathbf{M} \mid \mathbf{M} = \mathbf{M}^*\}$, i.e. is optimal, we also have that $\mathbf{M}\mathbf{\Gamma}_m = \begin{bmatrix} \mathbf{\Gamma}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix} \in \{\mathbf{M} \in \mathcal{S}_\mathbf{M} \mid \mathbf{M} = \mathbf{M}^*\}$, i.e. $\begin{bmatrix} \mathbf{\Gamma}_m \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix}$ is optimal. Then, since for $\mathbf{M}\mathbf{\Gamma}_m$ we have that $\mathcal{G}\left(\mathbf{M}\mathbf{\Gamma}_m\right) = \prod_{i=1}^m 1 + \frac{1}{\lambda_i\left(\hat{\mathbf{\Lambda}}_\mathbf{P}\right)}$, the proof is complete. □

APPENDIX E: PROOF OF PROPOSITION 3.3.1

First of all, we should reemphasize that the reason for setting a bound on the cumulative distribution function of $\|\beta\|^2$ is that this distribution function has a non-tractable characteristic because the random vector β has potentially non-identically and correlated distributed components. Then, our aim is to set a bound related to a function of β having i.i.d. components.

Now, consider

$$\frac{\|\hat{\mathbf{A}}\beta\|^2}{\|\beta\|^2} \leq \left|\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\right|, \tag{E.1}$$

which is true for all $\hat{\mathbf{A}} \in \mathbb{R}^{m\times m}$ (see [15]). Here, since $\mathbf{\Sigma}_\beta = \mathbf{\Lambda_G}\mathbf{U}_\mathbf{G}^T\mathbf{U}_\gamma\mathbf{\Lambda}_\gamma^2\mathbf{U}_\gamma^T\mathbf{U_G}\mathbf{\Lambda_G}$ by the construction of β (see Remark 3.3.1), if we let $\hat{\mathbf{A}} = \mathbf{\Lambda}_\gamma^{-1}\mathbf{U}_\gamma^T\mathbf{U_G}\mathbf{\Lambda}_\mathbf{G}^{-1}$ (that allows us to conclude that $\hat{\mathbf{A}}^T\hat{\mathbf{A}} > 0$ since $\hat{\mathbf{A}}$ is invertible, implying $|\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)| = \lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right) > 0$), we get $\mathrm{Cov}\left(\hat{\mathbf{A}}\beta\right) = \mathbf{I}_m$. So, $\hat{\mathbf{A}}\beta$ is a zero mean normal variable with the covariance matrix $\mathbf{I}_m$ (see [17]). Then by (E.1) we have

$$\begin{aligned}
\Pr_\beta\left[\beta^T\beta < \delta\right] &\leq \Pr_\beta\left[\|\beta\hat{\mathbf{A}}\|^2 < \lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\delta\right] &\text{(E.2)}\\
&= \mathcal{P}\left(\frac{m}{2}, \frac{\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right)\delta}{2}\right), &\text{(E.3)}
\end{aligned}$$

where (E.3) follow from the fact that the chi-squraed random variable $\|\hat{\mathbf{A}}\beta\|^2$, where $\hat{\mathbf{A}}\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$, has the cumulative distribution function $F_{\|\hat{\mathbf{A}}\beta\|^2}(x) = \mathcal{P}\left(\frac{m}{2}, \frac{x}{2}\right)$ (see [17]). Now, we can further simplify (E.3) to see the dependence of the cost on $\mathbf{T}$ in the following way,

$$\begin{aligned}
\lambda_{max}\left(\hat{\mathbf{A}}^T\hat{\mathbf{A}}\right) &= \lambda_{max}\left(\left(\mathbf{\Lambda_G}\mathbf{U}_\mathbf{G}^T\right)^{-1}\hat{\mathbf{A}}^T\hat{\mathbf{A}}\left(\mathbf{\Lambda_G}\mathbf{U}_\mathbf{G}^T\right)\right) &\text{(E.4)}\\
&= \lambda_{max}\left((\mathbf{\Sigma}_\gamma\mathbf{G})^{-1}\right)\\
&= \left(2\lambda_{min}\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{\Sigma}_{y\,|\,z}^{-1}\mathbf{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)^{-1}\right)\right)^{-1}, &\text{(E.5)}
\end{aligned}$$

where (E.4) follows from the properties of eigenvalues (since $\mathbf{\Lambda_G U_G}$ is invertible because $\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}$ in Remark 3.3.1 is shown to be positive definite in Appendix 3.2.1) and (E.5) follows from the positive definiteness of $\hat{\mathbf{A}}^T\hat{\mathbf{A}}$ (because $\mathbf{\Lambda_G}, \mathbf{\Lambda}_\gamma, \mathbf{U_G}, \mathbf{U}_\gamma$ are all non-singular) and from $\mathbf{\Sigma}_\gamma = 2\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T$ (by the properties of eigenvalues of positive definite matrices, see [15]). Then, since by (3.27) and (3.31) $P_\alpha = \Pr_\beta\left[\beta^T\beta < \delta\right]$, the proof is complete. □

APPENDIX F: PROOF OF THEOREM 3.3.1

We first let $\mathcal{H}(\mathbf{T}) = \lambda_{min}\left(\mathbf{T}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_{y\,|\,z}^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\right)$. Then by Theorem 3.1.1, we have

$$\mathcal{H}(\mathbf{T}) = \lambda_{min}\left(\mathbf{T}\boldsymbol{\Sigma}_x\left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_e - \boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\mathbf{T}\boldsymbol{\Sigma}_x\right)^{-1}\boldsymbol{\Sigma}_x\mathbf{T}^T\left(\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}^T\right)^{-1}\right). \tag{F.1}$$

Similar to the analysis in Appendix C, let $\boldsymbol{\Sigma}_x = \mathbf{F}\boldsymbol{\Lambda}^2\mathbf{F}^T$, where $\mathbf{F}$ is unitary and $\boldsymbol{\Lambda}$ is invertible diagonal (since $\boldsymbol{\Sigma}_x > 0$), and $\mathbf{K} \stackrel{\triangle}{=} \mathbf{T}\mathbf{F}\boldsymbol{\Lambda}$. This yields

$$\mathcal{H}(\mathbf{T}) = \lambda_{min}\left(\mathbf{K}\boldsymbol{\Lambda}\mathbf{F}^T\left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_e - \mathbf{F}\boldsymbol{\Lambda}\left(\mathbf{K}^T\left[\mathbf{K}\mathbf{K}^T\right]^{-1}\mathbf{K}\right)\boldsymbol{\Lambda}\mathbf{F}^T\right)^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{K}^T\left[\mathbf{K}\mathbf{K}^T\right]^{-1}\right) \tag{F.2}$$

If by SVD of $\mathbf{K}$ we write (where $\mathbf{K}$ is full-rank, and hence $\mathbf{U}_\mathbf{K}$ is unitary and $\boldsymbol{\Lambda}_\mathbf{K}$ is invertible diagonal) $\mathbf{K} = \mathbf{U}_\mathbf{K}\boldsymbol{\Lambda}_\mathbf{K}\mathbf{V}_\mathbf{K}^T$, $\mathcal{H}(\mathbf{T})$ reduces to

$$\begin{aligned}
\mathcal{H}(\mathbf{T}) &= \lambda_{min}\left(\mathbf{U}_\mathbf{K}\boldsymbol{\Lambda}_\mathbf{K}\mathbf{V}_\mathbf{K}^T\left(\mathbf{P} - \mathbf{V}_\mathbf{K}\mathbf{V}_\mathbf{K}^T\right)^{-1}\mathbf{V}_\mathbf{K}\boldsymbol{\Lambda}_\mathbf{K}^{-1}\mathbf{U}_\mathbf{K}^T\right) && \text{(F.3)}\\
&= \lambda_{min}\left(\mathbf{V}_\mathbf{K}^T\left(\mathbf{P} - \mathbf{V}_\mathbf{K}\mathbf{V}_\mathbf{K}^T\right)^{-1}\mathbf{V}_\mathbf{K}\right), && \text{(F.4)}\\
&= \lambda_{min}\left(\left(\mathbf{I}_m - \mathbf{V}_\mathbf{K}^T\mathbf{P}^{-1}\mathbf{V}_\mathbf{K}\right)^{-1} - \mathbf{I}_m\right), && \text{(F.5)}\\
&= \frac{1}{\lambda_{max}\left(\mathbf{I}_m - \mathbf{V}_\mathbf{K}^T\mathbf{P}^{-1}\mathbf{V}_\mathbf{K}\right)} - 1 && \text{(F.6)}
\end{aligned}$$

where $\mathbf{P} = \boldsymbol{\Lambda}^{-1}\mathbf{F}^T\left(\boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_e\right)\mathbf{F}\boldsymbol{\Lambda}^{-1}$, (F.4) follows from the properties of eigenvalues, (F.5) follows from (C.6) and (F.6) follows from the properties of eigenvalues and from the fact that $\mathbf{I}_m - \mathbf{V}_\mathbf{K}^T\mathbf{P}^{-1}\mathbf{V}_\mathbf{K} > 0$ with all eigenvalues satisfying $0 < \lambda_i < 1$, by Remark C.0.3. Here if we let $\mathbf{P} = \mathbf{U}_\mathbf{P}\boldsymbol{\Lambda}_\mathbf{P}\mathbf{U}_\mathbf{P}^T$, since $\mathbf{P}$ is symmetric, and define $\mathbf{M} \stackrel{\triangle}{=} \mathbf{U}_\mathbf{P}^T\mathbf{V}_\mathbf{K}$, we get

$$\mathcal{H}(\mathbf{T}) = \frac{1}{\lambda_{max}\left(\mathbf{M}^T\left(\mathbf{I}_n - \boldsymbol{\Lambda}_\mathbf{P}^{-1}\right)\mathbf{M}\right)} - 1 \tag{F.7}$$

Now, if we suppose that there exists $\mathbf{M}^*$ maximizing $\mathcal{H}(\mathbf{T})$ in (F.7), then since the construction of $\mathbf{M}$ for Proposition 3.2.2 and construction of $\mathbf{M}$ here are the same, by Proposition 3.2.2 we get the existence of $\mathbf{T}^*$ with the same construction in Appendix C (see that Remark C.0.5 still remains valid because we again have a cost function that is invariant to permutations, see [15]).

Moreover, the construction for a set of optimal $\mathbf{M}$s, i.e., construction of $\mathcal{M}$, in Proposition 3.2.3 is shown to satisfy that every $\mathbf{M} \in \mathcal{M}$ minimizes $\lambda_i\left(\mathbf{M}^T\left(\mathbf{I}_n - \mathbf{\Lambda}_{\mathbf{P}}^{-1}\right)\mathbf{M}\right)$ for all $i \in \{1, 2, ..., m\}$ by *Poincaré seperation theorem*, yielding that it particularly minimizes $\lambda_{max}\left(\mathbf{M}^T\left(\mathbf{I}_n - \mathbf{\Lambda}_{\mathbf{P}}^{-1}\right)\mathbf{M}\right)$ and hence maximizes $\mathcal{H}(\mathbf{T})$. Moreover, the minimum value is given by $\lambda_m\left(\mathbf{M}^T\left(\mathbf{I}_n - \mathbf{\Lambda}_{\mathbf{P}}^{-1}\right)\mathbf{M}\right) = \lambda_m\left(\hat{\mathbf{\Lambda}}_{\mathbf{P}}\right)$, where $\lambda_m\left(\hat{\mathbf{\Lambda}}_{\mathbf{P}}\right)$ indicates the m^{th} smallest eigenvalue of $\hat{\mathbf{\Lambda}}_{\mathbf{P}}$ given in Proposition 3.2.2. This yields the minimum value $\mathcal{P}\left(m/2, \delta/4\left(\frac{1}{\lambda_m\left(\hat{\mathbf{\Lambda}}_{\mathbf{P}}\right)} - 1\right)\right)$ for the cost function. Consequently, the set of transforms corresponding to this optimal set of $\mathbf{M}$s, i.e. the set $\mathcal{T}$, is a set of linear transforms in $\mathcal{S}_{\mathbf{T}}$ that are optimal in the sense of the bound on the cumulative distribution function of the conditional probability of error in (3.37). So we are done.

□

APPENDIX G: PROOF OF THEOREM 4.1.1

We first derive the probability of the error event of the detector with respect to the detection rule given by (4.1) as a function of $\mathbf{u} = J(i, \mathbf{x}_0, \mathbf{x}_1, \mathbf{T})$. First of all, if $\mathbf{z}_0 = \mathbf{z}_1$ then the probability of error, is exactly $1/2$ since there is nothing to separate $i = 0$ and $i = 1$ in terms of (4.1), regardless of $\mathbf{u}$. Hence for $\mathbf{z}_0 = \mathbf{z}_1$ there is nothing to optimize for $\mathbf{u}$, and any $\mathbf{u} \leq P$ a.s. satisfies (4.3), so let $\mathbf{u} = 0$ if $\mathbf{z}_0 = \mathbf{z}_1$. Now suppose $\mathbf{z}_0 \neq \mathbf{z}_1$ and that H_0 is the true hypothesis, i.e., $\mathbf{x}_0$ is transmitted and $\{\mathbf{z}_0, \mathbf{z}_1\}$ are given. Then by (4.1) we have

$$\begin{aligned}
\Pr\left[\text{error} \,\middle|\, H_0\right] &= \Pr\left[\|\mathbf{T}(\mathbf{x}_0 + \mathbf{u} + \mathbf{e}) - \mathbf{z}_0\|^2 > \|\mathbf{T}(\mathbf{x}_0 + \mathbf{u} + \mathbf{e}) - \mathbf{z}_1\|^2 \,\middle|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right] \\
&= \Pr\left[(\mathbf{z}_0 - \mathbf{z}_1)^T \mathbf{T}(\mathbf{u} + \mathbf{e}) < -\frac{\|\mathbf{z}_0 - \mathbf{z}_1\|^2}{2} \,\middle|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right] \qquad \text{(G.1)} \\
&= \Pr\left[a + b < c \,\middle|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right], \qquad \text{(G.2)}
\end{aligned}$$

where $a, b, c \in \mathbb{R}$ satisfy $a = (\mathbf{z}_0 - \mathbf{z}_1)^T \mathbf{T}\mathbf{u}, b = (\mathbf{z}_0 - \mathbf{z}_1)^T \mathbf{T}\mathbf{e}$ and $c = -\frac{\|\mathbf{z}_0 - \mathbf{z}_1\|^2}{2}$. Here since $\{\mathbf{z}_0, \mathbf{z}_1\}$ are given, by the properties of normal random variables (see [17]) and by Sec. 2.2, b is a zero mean normal random variable with a variance given by $(\mathbf{z}_0 - \mathbf{z}_1)^T \mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)$, which is nonzero since $\mathbf{z}_0 \neq \mathbf{z}_1$ and $\mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T > 0$ since $\mathbf{T}$ is full-rank and $\mathbf{\Sigma}_e > 0$ by assumption. Also $c \in \mathbb{R}$ is a constant and $a \in \mathbb{R}$ has a distribution induced by $\mathbf{u}$ (since $\mathbf{u}$ is potentially random). Then we have

$$\begin{aligned}
\Pr\left[\text{error} \,\middle|\, H_0\right] &= \mathrm{E}_{\{a,b\}}\left(\mathcal{X}_{A(a,b)}\right) \qquad \text{(G.3)} \\
&= \mathrm{E}_a\left(\mathrm{E}_{b|a}\left(\mathcal{X}_{A(a,b)}\right)\right) \qquad \text{(G.4)} \\
&= \mathrm{E}_a\left(\mathrm{Q}\left(\frac{a - c}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\|}\right)\right) \qquad \text{(G.5)}
\end{aligned}$$

where $\mathcal{X}_{A(a,b)}$ is the characteristic function of the set $A(a,b) = \{a+b \in \mathbb{R} | a+b < c\}$ and (G.5) follows from the fact that a and b are independent (since $\mathbf{u}$ and $\mathbf{e}$ are independent by assumption and linear functions are measurable) and $\Pr[b < c - a \,|\, a, \{\mathbf{z}_0, \mathbf{z}_1\}] = \mathrm{Q}\left(\frac{a-c}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\|}\right)$.

Then, since $a \geq -\|\mathbf{u}\|\|\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\| \geq -\sqrt{P}\|\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\|$ a.s. (by Cauchy-Schwarz inequality and peak power constraint) and Q-function is a strictly decreasing function of its argument, we get

$$\Pr\left[\text{error} \,\middle|\, H_0\right] \leq \mathrm{Q}\left(\frac{-\sqrt{P}\|\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\| - c}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T(\mathbf{z}_0 - \mathbf{z}_1)\|}\right). \tag{G.6}$$

The upper bound given by (G.6) is achieved if and only if $\mathbf{u} = -\sqrt{P}\frac{\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)}{\|\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)\|}$ a.s. (since both Cauchy-Schwarz and peak power inequalities are satisfied with equality if and only if $\mathbf{u} = -\sqrt{P}\frac{\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)}{\|\mathbf{T}^T(\mathbf{z}_0-\mathbf{z}_1)\|}$ a.s.). Moreover the case where H_1 is true is similar, yielding $\mathbf{u}^* = -\sqrt{P}\frac{\mathbf{T}^T(\mathbf{z}_1-\mathbf{z}_0)}{\|\mathbf{T}^T(\mathbf{z}_1-\mathbf{z}_0)\|}$ a.s. Then, since $\Pr\left[\text{error} \,\middle|\, \{\mathbf{z}_0, \mathbf{z}_1\}\right] = 1/2\Pr\left[\text{error} \,\middle|\, H_0\right] + 1/2\Pr\left[\text{error} \,\middle|\, H_1\right]$ where by (G.6) $\Pr\left[\text{error} \,\middle|\, H_0\right] = \Pr\left[\text{error} \,\middle|\, H_1\right]$, we get

$$\mathbf{u}^* = \begin{cases} -\sqrt{P}\frac{\mathbf{T}^T(\mathbf{z}_i-\mathbf{z}_j)}{\|\mathbf{T}^T(\mathbf{z}_i-\mathbf{z}_j)\|} \text{ a.s.} & \text{if } \mathbf{z}_i \neq \mathbf{z}_j \\ \mathbf{0} \text{ a.s.} & \text{if } \mathbf{z}_i = \mathbf{z}_j \end{cases} \tag{G.7}$$

where $\mathbf{u}^*$ is defined by (4.3). (G.7) clearly implies that any $\mathbf{u}$ that satisfies (4.4) also satisfies (4.3) (with restriction to equality everywhere, not only almost surely). Hence the proof is complete. □

APPENDIX H: PROOF OF PROPOSITION 4.2.1

We should not forget that, the cumulative distribution function of the Euclidean norm of a "correlated Gaussian" vector is not given in terms of known tractable functions. From this assertion, it is clear that the cumulative distribution function of the ratio of the Euclidean norms of such random vectors are not tractable. Then we follow the result in Remark 4.2.1 and step by step set upper bounds on P_{α,J^*}, which, as a result, yield a tractable form for the design of optimal $\mathbf{T}$. So we have

$$
\begin{aligned}
P_{\alpha,J^*} &= \Pr_\gamma \left[\frac{-\sqrt{P}\|\mathbf{T}^T\gamma\|}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|} + \frac{\|\gamma\|^2}{2\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|} < \mathrm{Q}^{-1}(\alpha) \,\middle|\, \gamma \neq 0 \right] && \text{(H.1)} \\
&\leq \Pr_\gamma \left[\frac{\|\gamma\|^2}{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|} < 2\left[\mathrm{Q}^{-1}(\alpha) + \sqrt{\frac{P}{\lambda_{min}(\mathbf{\Sigma}_e)}} \right] \right] && \text{(H.2)} \\
&\leq \Pr_\gamma \left[\|\gamma\|^2 < \beta^2 \lambda_{max}\left(\mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T\right) \right] && \text{(H.3)} \\
&\leq \Pr_\eta \left[\|\eta\|^2 < \frac{\beta^2 \lambda_{max}\left(\mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T\right)}{2\lambda_{min}\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)} \right] && \text{(H.4)} \\
&= \mathcal{P}\left(\frac{m}{2}, \frac{\beta^2 \lambda_{max}\left(\mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T\right)}{4\lambda_{min}\left(\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T\right)} \right), && \text{(H.5)}
\end{aligned}
$$

where $\beta = 2\left[\mathrm{Q}^{-1}(\alpha) + \sqrt{\frac{P}{\lambda_{min}(\mathbf{\Sigma}_e)}}\right]$ and $\eta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$. Here, (H.2) follows from the fact that $\frac{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|}{\|\mathbf{T}^T\gamma\|} \geq \sqrt{\lambda_{min}(\mathbf{\Sigma}_e)}$, (H.3) follows from $\frac{\|\mathbf{\Sigma}_e^{1/2}\mathbf{T}^T\gamma\|^2}{\|\gamma\|^2} \leq \lambda_{max}\left(\mathbf{T}\mathbf{\Sigma}_e\mathbf{T}^T\right)$ and (H.4) follows from $\frac{\|\mathbf{\Sigma}_\gamma^{-1/2}\gamma\|^2}{\|\gamma\|^2} \leq \lambda_{min}(\mathbf{\Sigma}_\gamma)$, $\mathbf{\Sigma}_\gamma = 2\mathbf{T}\mathbf{\Sigma}_x\mathbf{T}^T$ and $\eta = \mathbf{\Sigma}_\gamma^{-1/2}\gamma \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$. For the last equality, see [17]. □

REFERENCES

1. Shannon, C. E., "Channels With Side Information At The Transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, 1958.

2. Wyner, A., "On Source Coding With Side Information At The Decoder," *IEEE Transactions on Information Theory*, IT–21, pp. 294–300, 1975.

3. Wyner, A. and J. Ziv, "The Rate Distortion Function For Source Coding With Side Information At The Receiver," *IEEE Transactions on Information Theory*, IT-22, pp. 1–11, 1976.

4. Costa, M., "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439-441, May 1983.

5. Venkatesan, R., S. M. Koon, M. H. Jakubowski, P. Moulin, "Robust image hashing", in *Proc. IEEE Int. Conf. Image Processing, vol. 3,* pp. 664-666, 2000.

6. Mihcak, M. K. and R. Venkatesan, "A Perceptual Audio Hashing Algorithm: A Tool For Robust Audio Identification and Information Hiding," in *Proceedings of 4th International Information Hiding Workshop*, 2001.

7. Kozat, S. S., R. Venkatesan and M. K. Mihcak, "Robust Hashing via Matrix Invariances," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2004.

8. Monga, V. and M. K. Mihcak, "Robust and Secure Image Hashing via Non-Negative Matrix Factorizations," *IEEE Trans. Information Forensics and Security*, vol. 2, no. 3, pp. 376–390, Sep. 2007.

9. McEliece, R. J. and W. E. Stark, "An Information Theoretic Study of Communication In The Presence of Jamming," in *Proc. 1981 IEEE Intl. Conf. Commun.*,

Denver, CO, 1981.

10. Başar, T., "The Gaussian Test Channel With An Intelligent Jammer," *IEEE Transactions on Information Theory*, vol. IT–29, no. 1, Jan. 1983.

11. Başar, T., "Communication Games With Partially Soft Power Constraint," *Journal of Optimization Theory and Applications*, vol. 61, no. 3, June 1989.

12. Hegde, M. V., W. E. Stark, and D. Teneketzis, "On The Capacity of Channels With Unknown Interference," *IEEE Transactions on Information Theory*, vol. 35, no. 4, July 1989.

13. Mıhçak, M. K., Y. Altuğ and N. P. Ayerden, "On Minimax Optimal Linear Transforms for Detection with Side Information in Gaussian Setup," *IEEE Communications Letters*, vol. 12, no. 3, pp. 164–166, Mar. 2008.

14. Poor, H. V., *An Introduction to Signal Detection and Estimation*, Springer–Verlag, New York, 1988.

15. Horn, R. A. and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1999.

16. Loéve, M., *Probability Theory*, 2^{nd} ed., D. Van Nostrand Co., Inc., 1960.

17. Papoulis, A., *Probability, Random Variables and Stochastic Processes*, New York: McGraw Hill, 1965.

18. Moschovakis, Y. N., *Notes on Set Theory*, New York: Springer, 2006.