Channel Coding via Robust Optimization Part 1: The Single-User Channel

Chaithanya Bandi* Dimitris Bertsimas†

August 2015

Abstract

In this paper, we consider the optimal finite length (n) channel coding problem on a single-user Gaussian channel achieving an error probability bound of ϵ. We build on the ideas of Shannon [1948] and the paradigm of robust optimization (RO) to present novel optimization formulations of the channel coding problem on a single-user Gaussian, exponential and additive uniform noise channels. Solving these problems to optimality leads to upper and lower bounds on the channel capacity for finite code length n. As $n \to \infty$, the upper and lower bounds coincide and are equal to the Shannon capacity. The nature and computational complexity of the optimization problems depend on the noise distribution in the following way: (a) For Gaussian channels the optimization problems involve a rank minimization problem subject to semidefinite constraints, which we solve by iteratively solving semidefinite optimization problems; and (b) For exponential and additive uniform noise channels, the optimization problems are mixed integer linear optimization problems, which we solve using commercial solvers. Because of the size and complexity of these formulations, we do not solve them to provable optimality. Still we provide a feasible code that leads to a valid lower bound on channel capacity, but we only provide approximations for the upper bound. We report these computations for $n = 140$ for Gaussian channels and $n = 300$ for exponential channels.

1 The Channel Coding Problem

The central problem of communications is how to transmit information reliably through a noisy (and thus unreliable) communication channel. Shannon's "A mathematical theory of communication" published in 1948 marked the beginning of Information Theory. In this seminal paper he developed a framework to precisely define the intuitive notion of information, which in turn makes it possible to mathematically capture the notions of a communication channel and its capacity. Shannon showed that **(a)** there is an upper bound (the so called *channel capacity*) for the rate of reliable transmission of information, and **(b)** there exists a code that leads to a rate of transmission arbitrarily close to the channel capacity achieving probabilities of error arbitrarily close to zero. Since then, Shannon's approach has been used to characterize the fundamental limits of communication for various kinds of channels. However, the communication limits of many common channels

*Assistant Professor, Kellogg School of Management, Northwestern University, IL 60208, USA. Email: c-bandi@kellogg.northwestern.edu.

†Boeing Professor of Operations Research, co-director, Operations Research Center, Massachusetts Institute of Technology, E40-147, Cambridge, MA 02139, USA. Email: dbertsim@mit.edu.

such as the interference and the broadcast channels still remain unknown. Indeed a general theory for communication limits on networks of channels is still largely open. Techniques such as random encoding that are effective for single-user channels no longer allow us to characterize the capacity regions of complex channels.

The key difficulty in these problems is the analytical complexity of the optimization problems involved, which seek to maximize the number of symbols that can be transmitted over a channel while achieving a bounded probability of error. The probability of error for a given code involves an $n-$dimensional integral and is therefore computationally expensive, and optimization for such a code directly is quite complex. In this paper, we consider these optimization problems and use a robust optimization (RO) framework to reformulate these problems into optimization problems that are amenable to be solved by state of the art solvers. Solving these problems to optimality leads to upper and lower bounds on the channel capacity for finite code length n. As $n \to \infty$, the upper and lower bounds coincide and are equal to the Shannon capacity. Because of the size and complexity of these formulations, we do not solve them to provable optimality. Still we provide a feasible code that leads to a valid lower bound on channel capacity, but we only provide approximations for the upper bound. We report these computations for $n = 140$ for Gaussian channels and $n = 300$ for exponential channels.

In the present paper, which represents Part I of our work, we present the key steps in our approach and show how we reformulate the Gaussian channel coding problem as a rank minimization problem with semidefinite optimization constraints and the exponential and additive uniform noise channel coding problem as a mixed linear integer optimization problem. In Part II of our work (Bandi and Bertsimas [2015]), we address multi-user channels with interference and show that they can be reformulated in the same manner as in the single-user channel case.
We next briefly describe the philosophy as well as the key ingredients of our approach:

1. *Optimization formulation of Decoding Constraints:* By identifying that decoding is a robustness property, we formulate the probabilistic decoding constraints as robustness constraints of an optimization formulation. For single-user Gaussian channels, it is well known that the minimum distance decoder is an optimal decoder Cover and Thomas [2006], and we obtain robust quadratic constraints to represent the decoding property.
2. *Using typical sets as uncertainty sets:* We use the original idea of Shannon [1948], that for the purpose of computing the capacity and constructing the underlying code, it suffices to consider noise sequences that belong to the so called "typical set". We interpret these typical sets as uncertainty sets in a RO setting. Imposing that the decoding constraints found in Step 1, hold for all noise sequences in the typical set naturally leads to a RO formulation.
3. *Using binary optimization to count probabilities:* Since our objective is to solve the same problems, involving probabilistic primitives, that the information theory community has addressed over the years, we need to be able to measure the probability of error. In order to achieve this we need to count the frequency at which errors occur, which we accomplish by using binary optimization.
4. *Using semidefinite or binary optimization to model non-convexities:* We reformulate the underlying RO problem as either a mixed binary linear optimization problem (for non-Gaussian channels) or a non-convex quadratic optimization problem (for Gaussian channels). For the case of Gaussian channels, we reformulate the non-convex quadratic optimization problem as a rank minimization problem with semidefinite constraints. We then use the log-Det method developed in Fazell et al. [2003] to solve such problems as a sequence of semidefinite optimization problems.

1.1 Problem Definition and Notation

Throughout the paper, we denote scalar quantities by non-bold face symbols (e.g., $x \in \mathbb{R}$, $k \in \mathbb{N}$), vector quantities by boldface symbols (e.g., $\mathbf{x} \in \mathbb{R}^n, n > 1$), and matrices by uppercase boldface symbols (e.g., $\mathbf{A} \in \mathbb{R}^{n \times m}, n > 1, m > 1$). We denote scalar random variables as $\tilde{z}$ and vector random variables as $\tilde{\mathbf{z}}$. We use the notation $\tilde{\mathbf{z}} \sim N(\mathbf{0}, \sigma \cdot \boldsymbol{I})$ to denote that each component of $\tilde{\mathbf{z}}$ is normally distributed with mean 0 and standard deviation σ. For common information theoretic objects, we use the notation from Cover and Thomas [2006].

To make the paper self-contained, we define the notion of a communication channel and the related channel coding problem. Users send and receive messages over a *communication channel*. For example, in a single-user Gaussian channel, a sender transmits signal $\mathbf{x}_i \in \mathbb{R}^n$, but the receiver receives

$$\mathbf{y} = \mathbf{x}_i + \tilde{\mathbf{z}},$$

where the noise $\tilde{\mathbf{z}} \sim N(\mathbf{0}, \sigma \cdot \boldsymbol{I})$.

Given a communication channel, a sender seeks to transmit messages from a message set $\mathcal{M} = \{1, \ldots, M\}$ by coding the messages using codewords of length n, according to a code $\mathcal{C}$. The inputs of such a code $\mathcal{C}$ are:

(a) The length n of the codewords.

(b) The number $M = |\mathcal{M}| = 2^{nR}$ of codewords. The quantity $R = \log_2 M/n$ is called the rate of the code.

(c) The power constraint P of the sender.

(d) The noise standard deviation σ.

(e) The average probability of error $\epsilon > 0$ (see Eq. (1)) the user tolerates.

The outputs of $\mathcal{C}[n, R, P, \sigma, \epsilon]$ are:

(a) A code-book $\mathcal{B}$, which is a set of M codewords $\mathbf{x}_i$, $i = 1, \ldots, M$ satisfying $\|\mathbf{x}_i\|^2 \leq nP, \ \forall i$.

(b) A decoding function $g : \mathbb{R}^n \to \{1, 2, \ldots, M\}$ that maps each received word $\mathbf{y}$ to one of the codewords in $\mathcal{B}$, while satisfying the error-tolerance of ϵ. That is, for each $i = 1, \ldots, M$, we must have

$$\mathbb{P}[g(\mathbf{x}_i + \tilde{\mathbf{z}}) \neq i] \ \leq \ \epsilon, \tag{1}$$

with $\tilde{\mathbf{z}} \sim N(\mathbf{0}, \sigma \cdot \boldsymbol{I})$.

The finite capacity region of a single-user Gaussian channel $\mathcal{R}_n[P, \sigma, \epsilon]$ is the set of all rates R such that there exists a code $\mathcal{C}[n, R, P, \sigma, \epsilon]$. In the limit that $n \to \infty$ and $\epsilon \to 0$, the capacity region of a single-user Gaussian channel is called the asymptotic capacity region and is denoted by $\mathcal{R}[P, \sigma]$. In the next section, we present a brief review of the information theory literature organized around the results on different channels. We also review RO which is the key methodology that we use in this paper.

1.2 Relevant Literature

For a single-user communication channel, Shannon [1948] showed that there exists a maximum rate C associated with every communication channel, above which no reliable transmission is possible and below which there exists a code achieving small error probabilities. In particular, Shannon showed that the capacity C is given by

$$C = \sup I(X; Y),$$

where $I(X;Y)$ is the mutual information between the random variables X and Y, and the supremum is with respect to all possible input distributions of X. He showed that arbitrarily small probability of errors can be achieved by using random encoding with maximum likelihood decoding, whenever the rate of transmission is less than C. For the case of the Gaussian channel where the noise is normally distributed with mean 0 and standard deviation σ and the sender has a power constraint P, Shannon obtained that the asymptotic capacity of the channel is

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right) \text{ bits per channel use.} \tag{2}$$

For finite code length n, Polyanskiy et al. [2010] provide lower and upper bounds for the channel capacity. For a broad review of Information Theory we refer the reader to Verdú [1998], Verdú and McLaughlin [2000], Cover and Thomas [2006].

A Review of Binary Optimization

The modern theory of linear optimization (LO) started in the 1940s when George B. Dantzig proposed the simplex algorithm (Dantzig [1947]) for solving LO problems of the form

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ are given data and $\mathbf{x} \in \mathbb{R}^n$ is a vector of decision variables. The simplex algorithm proved practically efficient and it is, to this date, the main algorithm for solving LO problems. Today commercial solvers such as Cplex [2014] and Gurobi [2010] routinely solve problems with tens of millions of variables and constraints. For a review, see Bertsimas and Tsitsiklis [1997].

A natural and very relevant extension of LO is the class of mixed binary optimization (MBO) problems

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0},\ \mathbf{y} \in \{0,1\}^k, \end{aligned} \tag{3}$$

where in addition to the continuous variables $\mathbf{x}$, we also have binary variables $\mathbf{y}$. In the last six decades, significant progress has been made to solve MBO problems. Using commercial codes like Cplex [2014] and Gurobi [2010], we can routinely solve problems involving hundreds of thousands of binary variables and millions of continuous variables and constraints. In this paper, we utilize the ability of commercial solvers to solve large scale instances of problem (3). For a review of MBO, see Bertsimas and Weismantel [2005].

A Review of Robust Optimization

RO is one of the fastest growing areas of optimization in the last decade. It addresses the problem of optimization under uncertainty, in which the uncertainty model is not stochastic, but rather deterministic and set-based. RO models are typically polynomial-time solvable, but may lead to solutions that are too conservative. To alleviate conservatism, Ben-Tal and Nemirovski [2000,

1998, 1999], and El-Ghaoui and Lebret [1997], El-Ghaoui et al. [1998], proposed linear optimization models with ellipsoidal uncertainty sets, whose robust counterparts correspond to quadratic optimization problems. Bertsimas and Sim [2003, 2004] proposed RO models with polyhedral uncertainty sets that can model linear/integer variables, whose robust counterparts correspond to linear/integer optimization models. For a more thorough review we refer the reader to Bertsimas et al. [2011], Ben-Tal et al. [2009].

1.3 Contributions and Structure

The present paper is part of a broader research effort (Bandi and Bertsimas [2012]) to investigate a RO approach to classical problems with probabilistic primitives (like information theory). In our approach we replace the probabilistic primitives with uncertainty set based primitives and interpret their key properties as robustness properties. Our aim is to show that the resulting performance analysis or optimization questions become computationally tractable. We have implemented this program to single-class queueing networks in Bandi et al. [2015], pricing of multi-dimensional options in Bandi and Bertsimas [2014b], and in the problem of mechanism design Bandi and Bertsimas [2014a]. In the present paper, we consider the channel coding problem and present a robust optimization approach to this problem.

We begin by providing algorithms to compute the capacity region and to find optimal codes for the single-user Gaussian channel. We first present a RO approach to this channel that recovers the known asymptotic capacity and finds lower bounds for finite code length by solving a rank minimization subject to semidefinite constraints. We report computational results that show that the RO approach is computationally tractable for $n = 140$ for single-user Gaussian channels, $n = 300$ for single-user additive exponential noise channels. In Part II of this work, we extend this approach to the case of the two-user Gaussian interference channel, the two-user multi-access and broadcast Gaussian channels, and multi-user channels with exponentially distributed noise.

The structure of the paper is as follows. In Section 2, we present how decoding is a robustness property and the connection of typical sets and uncertainty sets. In Section 3, we introduce the RO approach for the single-user Gaussian channel. In Section 4, we examine how the resulting optimization problems depend on the nature of the probabilistic primitives and present mixed binary linear optimization problems to compute lower and upper bounds for the capacity region of single-user channels with exponentially distributed noise. In Section 5, we present some concluding remarks.

2 Robustness and Information Theory

In this section, we discuss how a robustness perspective can shed new light to information theory by interpreting **(a)** decoding a robustness property and **(b)** typical sets as uncertainty sets. We consider a single-user channel in which the noise is distributed according a probability density function (pdf) $f(\cdot)$.

2.1 Decoding as a Robustness Property

The Maximum Likelihood (ML) decoder is an optimal decoder for any single-user channel (see Cover and Thomas [2006]), that is, there always exists an optimal code which uses ML as the

decoding function. An ML decoder is characterized by the decoding function $g^{\mathrm{ML}}(\cdot)$ given by

$$g^{\mathrm{ML}}(\mathbf{y}) = \arg\max_{\mathbf{x}_i \in \mathcal{B}} \mathbb{P}[\mathbf{y}|\mathbf{x}_i \text{ was sent}] = \arg\max_{\mathbf{x}_i \in \mathcal{B}} \prod_{j=1}^{n} f(y_j - x_{ij}).$$

This allows us to formulate the coding problem as an optimization problem by restricting our attention to codes that are optimal with respect to the ML decoder. In particular, we constrain the codewords to satisfy the constraints

$$\prod_{j=1}^{n} f(x_{ij} + z_j - x_{i'j}) \;\leq\; \prod_{j=1}^{n} f(z_j), \;\forall i, i' \neq i, \forall \mathbf{z} \in \mathcal{U}_i, \tag{4}$$

where $\mathcal{U}_i$ is an appropriately chosen uncertainty set (see the discussion in Section 2.2) such that

$$\mathbb{P}[\tilde{\mathbf{z}} \in \mathcal{U}_i] \geq 1 - \epsilon., \tag{5}$$

and $\mathcal{U}_i$ could depend on the $\mathbf{x}_i$.
Note that Eqs. (4) are expressed in the language of robust optimization, which lead to Robust Optimization problems.

2.2 Typical Sets as Uncertainty Sets

Given a pdf $f(\cdot)$, Shannon [1948] in his study of the asymptotic capacity region of a communication channel, introduced the notion of a typical set $\mathcal{U}_\epsilon^f$ in order to capture the following two properties:

(a) $\mathbb{P}\left[\tilde{\mathbf{z}} \in \mathcal{U}_\epsilon^f\right] = 1 - \epsilon$, with $\epsilon \to 0$ as $n \to \infty$.

(b) The conditional pdf $h(\tilde{\mathbf{z}}) = f\left(\tilde{\mathbf{z}} \middle| \tilde{\mathbf{z}} \in \mathcal{U}_\epsilon^f\right)$ satisfies:

$$\left|\frac{1}{n}\log h(\tilde{\mathbf{z}}) + H_f\right| \leq \phi(\epsilon),$$

for some H_f (the differential entropy of the pdf $f(\cdot)$) and $\phi(\epsilon) \to 0$, as $n \to \infty$, where $\tilde{\mathbf{z}} = [\tilde{z}_1, \ldots, \tilde{z}_n]$ and $\tilde{z}_i \sim f(\cdot), \;\forall i = 1, \ldots, n$.

Property **(a)** means that the typical set has probability nearly one, while Property **(b)** means that all elements of the typical set are *nearly* equiprobable, see Cover and Thomas [2006]. We next show that the typical set of a pdf $f(\cdot)$ is given by

$$\mathcal{U}_\epsilon^f = \left\{ \mathbf{z} \left| -\Gamma_\epsilon^f \leq \frac{\sum_{i=1}^{n} \log f(z_i) + nH_f}{\sigma_f \sqrt{n}} \leq \Gamma_\epsilon^f. \right. \right\}, \tag{6}$$

where

$$H_f = -\int_{-\infty}^{\infty} f(x) \log f(x)\, \mathrm{d}x, \qquad \sigma_f^2 = \int_{-\infty}^{\infty} f(x) \left(\log f(x) + H_f\right)^2 \mathrm{d}x,$$

and Γ_ϵ^f is chosen such that

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} \log f(z_i) + nH_f\right| \leq \Gamma_\epsilon^f \cdot \sigma_f \sqrt{n}\right] = 1 - \epsilon. \tag{7}$$

Proposition 1. *For a pdf $f(\cdot)$, $\mathcal{U}_\epsilon^f$ defined in Eq. (6) satisfies*

(a) $\mathbb{P}\left[\tilde{\mathbf{z}} \notin \mathcal{U}_\epsilon^f\right] \leq \epsilon$.

(b) *The conditional pdf* $h(\tilde{\mathbf{z}}) = f\left(\tilde{\mathbf{z}} \middle| \tilde{\mathbf{z}} \in \mathcal{U}_\epsilon^f\right)$ *satisfies:* $\left|\frac{1}{n}\log h(\tilde{\mathbf{z}}) + H_f\right| \leq \phi(\epsilon)$, *with* $\phi(\epsilon) \to 0$, *as* $n \to \infty$.

The proof is ommitted as it is elementary.
The typical sets for the normal, exponential, uniform and binary distributions, are presented below.

Corollary 1. [Typical Sets for Normal, exponential, uniform and Binary Distributions]

(a) *The typical set for normally distributed i.i.d. random variables* $\tilde{z}_i \sim N(0,\sigma)$ *is given by*

$$\mathcal{U}_\epsilon^G = \left\{\mathbf{z} \,\middle|\, -\Gamma_\epsilon^G \leq \|\mathbf{z}\|^2 - n\sigma^2 \leq \Gamma_\epsilon^G\right\}, \tag{8}$$

(b) *The typical set for correlated normally distributed random variables* $\tilde{\mathbf{z}} \sim N(\mathbf{0},\mathbf{\Sigma})$ *is given by*

$$\mathcal{U}_\epsilon^{CG} = \left\{\mathbf{z} \,\middle|\, -\Gamma_\epsilon^{CG} \leq \|\mathbf{\Sigma^{-1}z}\|^2 - n \leq \Gamma_\epsilon^{CG}\right\}, \tag{9}$$

(c) *The typical set for exponentially distributed i.i.d. random variables* $\tilde{z}_i \sim Exp(\lambda)$ *is given by*

$$\mathcal{U}_\epsilon^E = \left\{\mathbf{z} \,\middle|\, \frac{n}{\lambda} - \frac{\sqrt{n}}{\lambda}\cdot\Gamma_\epsilon^E \leq \sum_{j=1}^n z_j \leq \frac{n}{\lambda} + \frac{\sqrt{n}}{\lambda}\cdot\Gamma_\epsilon^E,\ \mathbf{z} \geq \mathbf{0}\right\}, \tag{10}$$

(d) *The typical set for uniformly distributed i.i.d. random variables* $\tilde{z}_i \sim U[a,b]$ *is given by*

$$\mathcal{U}_\epsilon^U = \left\{\mathbf{z} \,\middle|\, \begin{array}{c} n\frac{a+b}{2} - \Gamma_\epsilon^U\sqrt{n} \leq \sum_{j=1}^n z_j \leq n\frac{a+b}{2} + \Gamma_\epsilon^U\sqrt{n}, \\ a \leq z_j \leq b,\ j = 1,\ldots,n, \end{array}\right\}, \tag{11}$$

(e) *The typical set for binary i.i.d. random variables* $\tilde{z}_i \sim Bin(p)$ *is given by*

$$\mathcal{U}_\epsilon^B = \left\{\mathbf{z} \,\middle|\, \begin{array}{c} np - \Gamma_\epsilon^B\sqrt{n} \leq \sum_{j=1}^n z_j \leq np + \Gamma_\epsilon^B\sqrt{n}, \\ z_j \in \{0,1\},\ j = 1,\ldots,n, \end{array}\right\}, \tag{12}$$

where $\Gamma_\epsilon^G,\ \Gamma_\epsilon^{CG},\ \Gamma_\epsilon^E,\ \Gamma_\epsilon^U,\ \Gamma_\epsilon^B$ *are chosen such that*

$$\mathbb{P}\left[\mathcal{U}_\epsilon^G\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^{CG}\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^E\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^U\right] = \mathbb{P}\left[\mathcal{U}_\epsilon^B\right] = 1-\epsilon. \tag{13}$$

single-

3 The Single-User Gaussian Channel

We consider a discrete time memoryless additive Gaussian channel, in which a single-user transmits a codeword $\mathbf{x}_i$ from a codebook $\mathcal{B}$. This codeword is transformed by the channel into $\mathbf{y} \in \mathbb{R}^n$ according to

$$\mathbf{y} = \mathbf{x}_i + \tilde{\mathbf{z}}_\mathbf{G},$$

where $\tilde{\mathbf{z}_G} \sim N(\mathbf{0}, \sigma\cdot\mathbf{I})$. The codewords are subject to an average power constraint P, that is, for any codeword $\mathbf{x}_i \in \mathcal{B}$ we require that

$$\|\mathbf{x}_i\|^2 \leq nP.$$

Suppose $\mathcal{B}$ consists of M codewords of length n and $\mathcal{M} = \{1,\ldots,M\}$. In what follows, we assume the code length is n unless otherwise mentioned, and we let $\tilde{\mathbf{z}}_G \sim N(\mathbf{0}, \sigma\cdot\mathbf{I})$ be the n-dimensional vector of i.i.d Normal random variables.

3.1 An Optimization Formulation of the Coding Problem

We begin by observing that the maximum likelihood decoder for a single user Gaussian channel reduces to the minimum distance decoder given by

$$g_0(\mathbf{y}) = \arg\min_{i \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}_i\|.$$

Using this decoder, we would want the codewords to satisfy the constraint

$$\|\mathbf{x}_i + \mathbf{z} - \mathbf{x}_{i'}\| \geq \|\mathbf{z}\| \quad \forall \mathbf{z} \in \mathcal{U}_\epsilon^i, \forall i, i' \neq i, \tag{14}$$

where $\mathcal{U}_\epsilon^i$ is a set of noise vectors with probability mass of $1 - \epsilon$. These constraints ensure that, if $\mathbf{x}_i$ was transmitted by the user and was received as

$$\mathbf{y} = \mathbf{x}_i + \mathbf{z} \text{ for some } \mathbf{z} \in \mathcal{U}_\epsilon^i,$$

then the distance between the received codeword $\mathbf{y}$ and any other codeword $\mathbf{x}_{i'}$ is greater than the distance between $\mathbf{y}$ and $\mathbf{x}_i$. Note that (14) is naturally expressed as a robustness property. The channel coding problem is, thus, given by

$$\begin{aligned} \max \quad & |\mathcal{M}| \\ \text{s.t.} \quad & \|\mathbf{x}_i + \mathbf{z} - \mathbf{x}_{i'}\| \geq \|\mathbf{z}\|, \quad \forall \mathbf{z} \in \mathcal{U}_\epsilon^i, \forall i, i' \in \mathcal{M}, i' \neq i, \\ & \|\mathbf{x}_i\|^2 \leq nP, \qquad \forall i \in \mathcal{M}. \end{aligned} \tag{15}$$

We will next present the optimal optimization formulation, and show how to model the decoding constraint in Eq. (14) with constraints in Eqs. (23) and (24), which uses the typical set for the Gaussian distribution given by Eq. (8). We begin by introducing a parameter ν that regulates the tradeoff between the accuracy of the computation of the finite capacity of the single-user Gaussian channel and the complexity of computing it; see also the discussion after Theorem 1.

Given inputs $n, R, P, \sigma, \epsilon, \nu$, we compute the following quantities:

1. The parameter γ_ϵ, which we choose so that

$$\mathbb{P}\left[\|\tilde{\mathbf{z}}_G\| \leq \gamma_\epsilon\right] \geq 1 - \epsilon. \tag{16}$$

2. The parameter T given by

$$T = \left(\frac{1+\nu}{\zeta\nu} \cdot \frac{\gamma_\epsilon}{\sqrt{n}}\right)^n, \text{ with } \zeta = \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - \epsilon\left(1 - \delta\left(\nu, n\right)\right)\right), \tag{17}$$

where $\Phi(\cdot)$ is the cdf of a standard normal and

$$\delta(\nu, n) = \exp\left(-n \cdot \frac{r - \log(1+r)}{2(1+3\nu)^2 \sigma^2}\right), \text{ with } r = \sigma^2\left((1+3\nu)^2 - (1+2\nu)^2\right),$$

and let $\mathcal{T} = \{1, \ldots, T\}$;

3. The parameter M_0 given by

$$M_0 = (1+\nu) \cdot \gamma_\epsilon. \tag{18}$$

4. The set of vectors

$$\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T\} \tag{19}$$

with $\|\mathbf{z}_t\| = M_0$, $t = 1, \ldots, T$, that are the deterministic equivalent of being uniformly distributed on the surface of n–dimensional sphere of radius M_0. The construction of such vectors has been studied under the umbrella of rate distortion theory (see Wyner [1967], Gray and Neuhoff [1998]). In Appendix B, we present an algorithm due to Lloyd [1982] to compute these vectors.

Code Construction

We next present an algorithm to construct codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ for the sender. Using the minimum distance decoder, we want to ensure that the average probability of error is at most ϵ, that is,

$$\frac{1}{M}\sum_{i\in\mathcal{M}}\mathbb{P}\left[g\left(\mathbf{y}\right)\neq i\,|m=i\right]\leq\epsilon. \tag{20}$$

In order to achieve this, we define the following "counting" variables $\{v_{it}\}_{i\in\mathcal{M},t\in\mathcal{T}}$:

$$v_{it}=\begin{cases}1, & \text{if } \|\mathbf{x}_i+\mathbf{z}_t-\mathbf{x}_{i'}\|\geq\|\mathbf{z}_t\|\,,\ \forall i'\in\mathcal{M},\\ 0, & \text{otherwise.}\end{cases} \tag{21}$$

The *Encoding Algorithm* is, thus, given by the feasibility problem:

$$\|\mathbf{x}_i\|^2\leq nP, \qquad \forall i\in\mathcal{M}, \tag{22}$$

$$\|\mathbf{x}_i-\mathbf{x}_k+\mathbf{z}_t\|+\left(1-v_{it}\right)M_0\geq\|\mathbf{z}_t\|\,, \quad \forall t\in\mathcal{T},\ \forall i,k\in\mathcal{M},\ k\neq i, \tag{23}$$

$$\sum_{t=1}^{T}v_{it}\geq\left(1-\epsilon\right)T, \qquad \forall i\in\mathcal{M}, \tag{24}$$

$$\|\mathbf{x}_i-\mathbf{x}_k\|\geq 2\zeta\sqrt{n}, \qquad \forall i,k\in\mathcal{M},\ k\neq i, \tag{25}$$

$$v_{it}\in\{0,1\}, \qquad \forall i\in\mathcal{M},t\in\mathcal{T}, \tag{26}$$

We next explain each of the constraints in the *Encoding Algorithm*:

1. The constraints (22) impose power constraints on the codewords.
2. The constraints (23) implement the decoding rule for noise vector $\mathbf{z}_t$.
3. The constraints (24) imposes that the decoding constraint needs to be valid for at least $(1-\epsilon)T$ of the $\mathbf{z}_t's$. In other words, a fraction $\epsilon\cdot T$ of noise vectors $\mathbf{z}_t$'s are not constrained to satisfy the minimum distance property.
4. The constraints (25) ensures that the codewords are separated by a certain minimum distance to obtain a decoding error probability of at most ϵ.

Semidefinite Programming Reformulation

We next reformulate the feasibility problem (22)-(26) as a semidefinite optimization problem with rank constraints. We do this in two steps: (a) reformulate constraints (22)-(26) into quadratic constraints by using the Proposition 2, and (b) reformulate quadratic constraints into SDO constraints by using Proposition 3.

Proposition 2.

(a) *Constraint (23) is equivalent to the constraint*

$$\|\mathbf{x}_i-\mathbf{x}_k\|^2+M_0^2\left(1-v_{it}\right)\geq 2\left(\mathbf{x}_k-\mathbf{x}_i\right)'\mathbf{z}_t, \forall t\in\mathcal{T},\ \forall i,k\in\mathcal{M},\ k\neq i.$$

(b) *Constraint $v_{it}\in\{0,1\}$ is equivalent to the constraint $v_{it}^2=v_{it}$.*

Proof. **(a)** When $v_{it} = 1$,

$$\begin{aligned} & \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| + (1 - v_{it})\, M_0 \geq \|\mathbf{z}_t\| \\ \iff & \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\|^2 \geq \|\mathbf{z}_t\|^2 \\ \iff & \|\mathbf{x}_i - \mathbf{x}_k\|^2 \geq 2\,(\mathbf{x}_k - \mathbf{x}_i)'\,\mathbf{z}_t. \end{aligned}$$

On the other hand, when $v_{it} = 0$,

$$\begin{aligned} & \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| + (1 - v_{it})\, M_0 \geq \|\mathbf{z}_t\| \\ \iff & \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \geq \|\mathbf{z}_t\| - M_0 \\ \iff & \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\|^2 \geq \|\mathbf{z}_t\|^2 + M_0^2 - 2M_0\,\|\mathbf{z}_t\| \\ \iff & \|\mathbf{x}_i - \mathbf{x}_k\|^2 + M_0^2 \geq 2\,(\mathbf{x}_k - \mathbf{x}_i)'\,\mathbf{z}_t, \end{aligned}$$

where the last equivalence follows from $\|\mathbf{z}_t\| = M_0$. Therefore,

$$\left\{\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| + (1 - v_{it})\, M_0 \geq \|\mathbf{z}_t\|\right\} \iff \left\{\|\mathbf{x}_i - \mathbf{x}_k\|^2 + M_0^2\,(1 - v_{it}) \geq 2\,(\mathbf{x}_k - \mathbf{x}_i)'\,\mathbf{z}_t\right\}.$$

(b) We have that $v_{it} \in \{0, 1\}$ if and only if $v_{it}^2 = v_{it}$. □

Using Proposition 2, we convert the feasibility problem (22)-(26) to a non-convex quadratic optimization problem. Let $\mathcal{K} = \{1, \ldots, K\}$.

Proposition 3. *The set of quadratic, possibly non-convex, constraints*

$$f_k(\mathbf{y}) = \mathbf{y}'\mathbf{A}_k\mathbf{y} + 2\mathbf{b}_k'\mathbf{y} + c_k \leq 0, \;\; \forall k \in \mathcal{K}. \tag{27}$$

is equivalent to the semidefinite optimization problem

$$\begin{aligned} \tilde{\mathbf{A}}_k \bullet \mathbf{Y} \leq 0, \quad \forall k \in \mathcal{K}, \\ Y_{11} = 1, \;\; \mathbf{Y} \succeq \mathbf{0}, \;\; \mathit{rank}\,(\mathbf{Y}) = 1, \end{aligned} \tag{28}$$

where

$$\mathbf{Y} = \begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} (1, \mathbf{y}'), \quad \tilde{\mathbf{A}}_k = \begin{pmatrix} c_k & \mathbf{b}_k \\ \mathbf{b}_k & \mathbf{A}_k \end{pmatrix}.$$

Proof. The quadratic function $f_k(\cdot)$ can be written as

$$f_k(\mathbf{y}) = (1, \mathbf{y}') \begin{pmatrix} c_k & \mathbf{b}_k \\ \mathbf{b}_k & \mathbf{A}_k \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} = \tilde{\mathbf{A}}_k \bullet \mathbf{Y},$$

where $\mathbf{Y} = \begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} (1, \mathbf{y}')$, and $\tilde{\mathbf{A}}_k = \begin{pmatrix} c_k & \mathbf{b}_k \\ \mathbf{b}_k & \mathbf{A}_k \end{pmatrix}$. Clearly,

$$Y_{11} = 1, \; \mathbf{Y} \succeq \mathbf{0}, \text{ and } \operatorname{rank}(\mathbf{Y}) = 1.$$

In addition, $\tilde{\mathbf{A}}_k \bullet \mathbf{Y} = f_k(\mathbf{y}) \leq 0, \;\; \forall k \in \mathcal{K}$.

On the other hand, given a feasible solution $\mathbf{Y}$ to (28), because $\operatorname{rank}(\mathbf{Y}) = 1$ and $Y_{11} = 1$, there exists a vector $\mathbf{y}$ such that

$$\mathbf{Y} = \begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} (1, \mathbf{y}'),$$

and clearly $\mathbf{y}$ is feasible to (27). □

Using Propositions 2 and 3 we next show that the feasibility problem (22)-(26) is equivalent to checking whether the optimal solution value of the semidefinite optimization problem (29) is equal to one.

$$\begin{aligned} \min \quad & \operatorname{rank}(\mathbf{Y}) \\ \text{s.t.} \quad & \mathbf{A}_i \bullet \mathbf{Y} \leq 0, \quad \forall i \in \mathcal{M}, \\ & \mathbf{B}_{ikt} \bullet \mathbf{Y} \leq 0, \quad \forall t \in \mathcal{T},\ \forall i, k \in \mathcal{M},\ k \neq i, \\ & \mathbf{C}_i \bullet \mathbf{Y} \leq 0, \quad \forall i \in \mathcal{M}, \\ & \mathbf{E}_{ik} \bullet \mathbf{Y} \leq 0, \quad \forall i, k \neq i \in \mathcal{M}, \\ & \mathbf{D}_{it} \bullet \mathbf{Y} = 0, \quad \forall i \in \mathcal{M}, t \in \mathcal{T}, \\ & \mathbf{Y} \succeq \mathbf{0}, \end{aligned} \tag{29}$$

where

$$\mathbf{A}_i(p,q) = \begin{cases} -nP, & \text{if } p = 1,\ q = 1, \\ 0, & \text{if } p = 1,\ q > 1, \\ 0, & \text{if } p > 1,\ q = 1, \\ 1, & \text{if } \forall p = q = (i-1)n+1, \ldots, in+1, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{C}_i(p,q) = \begin{cases} (1-\epsilon)\,T, & \text{if } p = 1,\ q = 1, \\ -1, & \text{if } \forall p = q = n^2+1+(i-1)T+t,\ \forall t = 1, \ldots, T, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{D}_{it}(p,q) = \begin{cases} 0, & \text{if } p = 1,\ q = 1, \\ -\frac{1}{2}, & \text{if } p = 1,\ q = n^2+1+(i-1)T+t, \\ -\frac{1}{2}, & \text{if } q = 1,\ p = n^2+1+(i-1)T+t, \\ 1, & \text{if } \forall p = q = n^2+1+(i-1)T+t, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{E}_{ik}(p,q) = \begin{cases} 4n\zeta^2, & \text{if } p = 1,\ q = 1, \\ 0, & \text{if } p = 1,\ q > 1, \\ 0, & \text{if } p > 1,\ q = 1, \\ 1, & \text{if } \forall q > 1, p = (i-1)n+1, \ldots, in+1, \\ -1, & \text{if } \forall p > 1, q = (k-1)n+1, \ldots, kn+1, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{B}_{ikt}(p,q) = \begin{cases} -M_0^2, & \text{if } p=1,\ q=1, \\ z_{tr}, & \text{if } p=1,\ q=(k-1)n+1+r,\ \forall r=1,\ldots,n, \\ z_{tr}, & \text{if } q=1,\ p=(k-1)n+1+r,\ \forall r=1,\ldots,n, \\ -z_{tr}, & \text{if } p=1,\ q=(i-1)n+1+r,\ \forall r=1,\ldots,n, \\ -z_{tr}, & \text{if } q=1,\ p=(i-1)n+1+r,\ \forall r=1,\ldots,n, \\ -1, & \text{if } \forall p=q=(i-1)n+1,\ldots,in+1, \\ -1, & \text{if } \forall p=q=(k-1)n+1,\ldots,kn+1, \\ 1, & \text{if } q=(k-1)n+1+r,\ p=(i-1)n+1+r,\ \forall r=1,\ldots,n, \\ 1, & \text{if } q=(i-1)n+1+r,\ p=(k-1)n+1+r,\ \forall r=1,\ldots,n, \\ \frac{M_0^2}{2} & \text{if } p=1,\ q=n^2+1+(i-1)T+t, \\ \frac{M_0^2}{2} & \text{if } q=1,\ p=n^2+1+(i-1)T+t, \\ 0, & \text{otherwise.} \end{cases}$$

We summarize the discussion by presenting the following algorithm which checks if a given rate R and an error probability ϵ are achievable by a sender with power P on a channel with noise variance σ using codewords of length n. As discussed before, the accuracy of the algorithm is based on a parameter ν.

Algorithm 1. ***Achievable rates on a Single-User Gaussian Channel***
Input: *Code parameters n, R, ϵ; Channel parameters P, σ; and accuracy parameter ν.*
Output: *Codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$, if the rate R is achievable.*
Algorithm:

1. *Solve the rank minimization semidefinite optimization problem (29) to compute r^*, codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$, and auxiliary binary variables $\{v_{it}\}_{t\in\mathcal{T}}$.*
2. *When $r^*=1$, then declare that the rate R is achievable using the codebook $\mathcal{B}=\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ and the minimum distance decoding function, achieving a decoding error probability of 2ϵ.*
3. *If $r^*\geq 2$, then declare that the rate R cannot be achieved on a single-user Gaussian channel with noise standard deviation $(1+3\nu)\sigma$ with probability of error less than or equal to $\epsilon(1-\delta(\nu,n))$, where*

$$\delta(\nu,n) = \exp\left(-n\cdot\frac{r-\log(1+r)}{2(1+3\nu)^2\sigma^2}\right),\ \text{with } r=\sigma^2\left((1+3\nu)^2-(1+2\nu)^2\right).$$

When a particular value of $\underline{R}$ leads to $r^*=1$, $\underline{R}$ represents a lower bound for the channel capacity. We also obtain a code that achieves the rate $\underline{R}$. When a particular value of $\overline{R}$ leads to $r^*\geq 2$, $\overline{R}$ represents an upper bound for the channel capacity. In Theorem 1, we show how we use Algorithm 1 to deduce lower and upper bounds on the capacity of a single-user Gaussian channel.

We next present the main result of this paper.

Theorem 1. *Let $\mathcal{R}_n[P,\sigma,\epsilon]$ is the set of all achievable rates on a single-user Gaussian channel with power P, standard deviation of the noise σ and maximum decoding error probability of ϵ. Consider the optimization problem (29) constructed for parameters $n, R, P, \sigma, \epsilon$ and let r^* and $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ be an optimal solution, then*

(a) *If* $r^* = 1$*, then* $R \in \mathcal{R}_n[P, \sigma, 2\epsilon]$*, that is the rate* R *is achievable using the codebook* $\mathcal{B} = \{\mathbf{x}_i\}_{i \in \mathcal{M}}$ *and the minimum distance decoding function, achieving a maximum decoding error probability of* 2ϵ*.*

(b) *If* $r^* \geq 2$*, then* $R \notin \mathcal{R}_n[P, (1+3\nu)\sigma, \bar{\epsilon} = \epsilon(1-\delta(\nu, n))]$*, which means that the rate* R *cannot be achieved on a channel with a noise variance* $(1+3\nu)\sigma$*, where*

$$\delta(\nu, n) = \exp\left(-n \cdot \frac{r - \log(1+r)}{2(1+3\nu)^2 \sigma^2}\right), \text{ with } r = \sigma^2\left((1+3\nu)^2 - (1+2\nu)^2\right).$$

Discussion
Theorem 1(a) indicates that for values of $(n, \epsilon/2, P, \sigma, \nu, R)$, if $r^* = 1$, then the rate R is achievable, and thus such an R provides an lower bound of the capacity $\mathcal{R}_n[P, \sigma, \epsilon]$.
Theorem 1(b) indicates that for values $(n, \epsilon/(1-\delta(\nu, n)), P, \sigma/(1+3\nu), \nu, R)$, if $r^* \geq 2$, then the rate R is not achievable, and thus such an R provides an upper bound on the capacity $\mathcal{R}_n[P, \sigma, \epsilon]$. In this way, Algorithm 1 provides upper and lower bounds for the capacity of a single-user Gaussian channel for finite n. In the limit of $\nu, \epsilon \to 0$ and $n \to \infty$ the lower and upper bounds are tight. So, in principle our approach provides valid upper and lower bounds. In numerical implementations, however, we do not solve problem (29) to provable optimality due to its size and complexity. When we find $r^* = 1$, we still provide a valid lower bound on channel capacity, but when we report $r^* \geq 2$, we do not have a guarantee as we have not solved problem (29) to provable optimality. In this way, the upper bound we report can only be seen as an approximation.

3.2 Proof of Theorem 1

We present the proof of Theorem 1 in this section. Before we proceed, we establish the following notation for this section. We let $\mathscr{E}_i[\tilde{\mathbf{z}}]$ denote the event that a decoding error occurs when message i is sent on the channel and noise vector $\tilde{\mathbf{z}}$ is realized, that is,

$$\mathscr{E}_i[\tilde{\mathbf{z}}] = \{\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}\| \leq \|\tilde{\mathbf{z}}\|\}.$$

Let $\mathbf{1}\{\mathscr{E}_i[\tilde{\mathbf{z}}]\}$ denote the indicator random variable corresponding to $\mathscr{E}_i[\tilde{\mathbf{z}}]$. Furthermore, let $\mathcal{S}_n(r)$ be the n–dimensional shell of radius r given by

$$\mathcal{S}_n(r) = \left\{\mathbf{s} \in \mathbb{R}^n \middle| \|\mathbf{s}\| = r\right\}, \tag{30}$$

and let $\tilde{\mathbf{s}}(r)$ denote a vector chosen uniformly at random in $\mathcal{S}_n(r)$. We next review the following two results from the literature which will be integral to the proof of correctness of Algorithm 1. Wyner [1967] showed that a large collection of uniformly random points on a sphere can be used as a good approximation for all the points on the sphere, as given in Proposition 4.

Proposition 4 (Wyner [1967])**.** *Let* $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N\}$ *be a Voronoi tessellation on* $\mathcal{S}_n(\gamma)$*. Then,* $\forall \mathbf{s} \in \mathcal{S}_n(\gamma)$*,* $\exists \mathbf{a}_i \in \mathcal{A}$ *such that* $\|\mathbf{s} - \mathbf{a}_i\| \leq \gamma / N^{1/n}$*.*

The following result lists two important properties of a vector of independent normal random variables.

Proposition 5 (Cover and Thomas [2006])**.** *Let* $\tilde{\mathbf{z}}_G \sim N(\mathbf{0}, \sigma \cdot \mathbf{I})$*.*

(a) (Bernstein's inequality) *The vectors* $\tilde{\mathbf{z}}_G$ *are concentrated in a thin shell of radius* $\sigma\sqrt{n}$*, that is,*

$$\mathbb{P}\left[\frac{1}{n}\left\|\tilde{\mathbf{z}}_G\right\|^2 > \sigma^2 - r\right] \geq 1 - \exp\left(-n \cdot \frac{r - \log(1+r)}{2\sigma^2}\right).$$

(b) (Spherical symmetry) *The random vector* $\tilde{\boldsymbol{u}} = \tilde{\mathbf{z}}_G / \left\|\tilde{\mathbf{z}}_G\right\|$ *is distributed uniformly in* $\mathcal{S}_n(1)$*.*

(c) *Let* $\tilde{d}$ *be a random variable distributed identically to the norm of* $\tilde{\mathbf{z}}_G$*, that is,* $\tilde{d} \sim \|\tilde{\mathbf{z}}_G\|$*. Then,* $\tilde{\mathbf{z}}_G \sim \tilde{d} \cdot \tilde{\mathbf{s}}_n(1)$*, where* $\tilde{\mathbf{s}}(1)$ *denote a vector chosen uniformly at random in* $\mathcal{S}_n(1)$*.*

We next present a series of propositions that form the components of the proof of Theorem 1. In Proposition 6, we present some of the geometric properties we need.

Proposition 6. *Let* $r_A, r_B, \gamma > 0$*. Let* A*,* B *and* C *be three single-user channels with noise vectors* $\tilde{\boldsymbol{u}}_A, \tilde{\boldsymbol{u}}_B, \tilde{\mathbf{z}}_C$ *where* $\tilde{\boldsymbol{u}}_A, \tilde{\boldsymbol{u}}_B$ *are distributed uniformly in* $\mathcal{S}_n(r_A)$ *and* $\mathcal{S}_n(r_B)$*,* respectively, and $\tilde{\boldsymbol{z}}_C \sim N(\mathbf{0}, \gamma \cdot \boldsymbol{e})$. Let $\{\mathbf{x}_i\}_{i \in \mathcal{M}}$ be the set of codewords. Then,

(a) If $r_A \geq r_B$, then $\mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_A\| \leq r_A\right] \geq \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_B\| \leq r_B\right]$;

(b) $\mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_A\| \leq r_A\right] \geq \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{z}}_C\| \leq \|\tilde{\boldsymbol{z}}_C\| \,\middle|\, \|\tilde{\boldsymbol{z}}_C\| \leq r_A\right]$;

(c) $\mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_A\| \leq r_A\right] \leq \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{z}}_C\| \leq \|\tilde{\boldsymbol{z}}_C\| \,\middle|\, \|\tilde{\boldsymbol{z}}_C\| > r_A\right]$.

In Proposition 7, we show that a code that is "good" with respect to Gaussian noise is also "good" for uniform noise.

Proposition 7. *Let* $\mathcal{C}$ *be a code with codebook* $\mathcal{B}$ *and a minimum distance decoding function. Consider two noises* $\tilde{\mathbf{z}}_G \sim N(\mathbf{0}, \sigma \cdot \mathbf{I})$ *and* $\tilde{\mathbf{z}}_U$ *uniformly distributed on* $\mathcal{S}_n(\overline{\sigma}\sqrt{n})$ *with* $\overline{\sigma} < \sigma$*. If* $\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\right] \leq \epsilon$*, then*

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_U\right]\right] \leq \frac{\epsilon}{1 - \exp(-n\beta)}, \text{ with } \beta = \frac{\sigma^2 - \overline{\sigma}^2 - \log(1 + \sigma^2 - \overline{\sigma}^2)}{2\sigma^2}.$$

We next examine certain properties of optimal solutions of (29). Let $\mathcal{W}$ be the set of all scaled noise vectors defined by $\mathcal{W} = \left\{\mathbf{w}_t | \mathbf{w}_t = \frac{1}{1+\nu}\mathbf{z}_t, \ \forall t \in \mathcal{T}\right\}$, and $\tau(\mathbf{s}) = \arg\min_{t \in \mathcal{T}} \|\mathbf{s} - \mathbf{w}_t\|$.

Proposition 8. *Let* $\epsilon > 0$ *and* γ_ϵ *be as in (16). Let* $\{r^*, \mathbf{x}_i, v_{it}\}$ *be an optimal solution of (29). If* $r^* = 1$*, then for* $\tilde{\mathbf{s}}$ *uniformly distributed on* $\mathcal{S}_n(\gamma_\epsilon)$*, we have*

(a) $\mathbb{P}\left[\forall k \neq i : \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{s}}\| \geq \gamma_\epsilon\right] \geq \mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\right]$,

(b) $\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\right] \geq 1 - \epsilon$.

All the proofs are presented in Appendix A. We next present the proof of Theorem 1.

Proof of Theorem 1.
In Part (a), we have $r^* = 1$ and therefore, we compute the codewords $\{\mathbf{x}_i\}$ that we then transmit on the channel. We then show that the probability of error when these codewords are transmitted on a single-user Gaussian channel with noise standard deviation σ is bounded by 2ϵ. In Part (b), when $r^* = 2$, we cannot compute feasible codewords $\{\mathbf{x}_i\}$, and we then show that rate R cannot be achieved on a channel with noise standard deviation $(1+3\nu)\sigma$ achieving an error probability $\bar{\epsilon}$. That is we show that if the rate R could be achieved on a channel with noise standard deviation $(1+3\nu)\sigma$ achieving an error probability $\bar{\epsilon}$, then it would have been accepted by Algorithm 1.

We begin by a proof of Part (a).

(a) Let $\tilde{\mathbf{z}}_G$ be an n-dimensional Gaussian noise with standard deviation σ, that is, $\tilde{\mathbf{z}}_G \sim N(\mathbf{0}, \sigma \cdot \mathbf{I})$, and let $f_G(\cdot)$ be the probability density function defined as

$$f_G(x) = \frac{\mathrm{d}P\left[\|\tilde{\mathbf{z}}_G\| \leq x\right]}{\mathrm{d}x}.$$

We next calculate the probability of error when a codeword i is sent on the channel:

$$\begin{aligned}
\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_G\right]\right] =& \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq \|\tilde{\mathbf{z}}_G\|\right] \\
=& \int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq \|\tilde{\mathbf{z}}_G\| \mid \|\tilde{\mathbf{z}}_G\| = c\right] \cdot f_G(c)\,\mathrm{d}c \\
&+ \int_{\gamma_\epsilon}^{\infty} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq \|\tilde{\mathbf{z}}_G\| \mid \|\tilde{\mathbf{z}}_G\| = c\right] \cdot f_G(c)\,\mathrm{d}c.
\end{aligned}$$

We bound the second term as follows :

$$\begin{aligned}
&\int_{\gamma_\epsilon}^{\infty} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq \|\tilde{\mathbf{z}}_G\| \mid \|\tilde{\mathbf{z}}_G\| = c\right] \cdot f_G(c)\,\mathrm{d}c \\
&\quad \leq \int_{\gamma_\epsilon}^{\infty} f_G(c)\,\mathrm{d}c = \mathbb{P}\left[\tilde{\mathbf{z}}_G \notin \mathcal{S}_n(\gamma_\epsilon)\right] \leq \epsilon,
\end{aligned} \tag{31}$$

which follows from the definition of γ_ϵ in (16).

We bound the first term as follows:

$$\begin{aligned}
&\int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq \|\tilde{\mathbf{z}}_G\| \mid \|\tilde{\mathbf{z}}_G\| = c\right] \cdot f_G(c)\,\mathrm{d}c \\
&\quad = \int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}_G\| \leq c \mid \|\tilde{\mathbf{z}}_G\| = c\right] \cdot f_G(c)\,\mathrm{d}c \\
&\quad = \int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{s}}\,(c)\| \leq c\right] \cdot f_G(c)\,\mathrm{d}c,
\end{aligned}$$

which follows from Prop. 5(c) by observing that conditioned on $\|\tilde{\mathbf{z}}_G\| = c$, the distribution of $\tilde{\mathbf{z}}_G$ is identical to $\tilde{\mathbf{s}}\,(c)$ which is a vector uniformly distributed on a $n-$ball of radius c. We next have that

$$\begin{aligned}
&\int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{s}}\,(c)\| \leq c\right] \cdot f_G(c)\,\mathrm{d}c \\
&\quad \leq \int_0^{\gamma_\epsilon} \mathbb{P}\left[\exists k \neq i \,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{s}}\,(\gamma_\epsilon)\| \leq \gamma_\epsilon\right] \cdot f_G(c)\,\mathrm{d}c \quad \text{(from Prop. 6(a) because } c \leq \gamma_\epsilon) \\
&\quad \leq \int_0^{\gamma_\epsilon} \left(1 - \mathbb{P}\left[v_{i\tau_{\tilde{\mathbf{s}}}(\gamma_\epsilon)} = 1\right]\right) \cdot f_G(c)\,\mathrm{d}c \qquad \text{(from Prop. 8(a))} \\
&\quad \leq \int_0^{\gamma_\epsilon} (1 - (1 - \epsilon)) \cdot f_G(c)\,\mathrm{d}c \qquad \text{(from Prop. 8(b))} \\
&\quad \leq \epsilon \cdot \int_0^{\gamma_\epsilon} f_G(c)\,\mathrm{d}c \leq \epsilon.
\end{aligned} \tag{32}$$

From (31) and (32), we have $\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_G\right]\right] \leq 2\epsilon$, and the Part (a) of theorem follows.

(b) To prove Part(b), we prove its contra-positive, that is, we show that if we choose a rate $R \in \mathcal{R}_n\left(P, (1+3\nu)\,\sigma, \overline{\epsilon}\right)$, and then there exists a code that is feasible to Constraints (22)-(26).

Consider a rate R such that $R \in \mathcal{R}_n\left(P, (1+3\nu)\,\sigma, \overline{\epsilon}\right)$, then by definition, there must exist a code $\mathcal{C} = \{\mathbf{x}_i\}_{i\in\mathcal{M}}$ that has an error-probability of $\overline{\epsilon}$ on the channel with Gaussian noise $\tilde{\mathbf{f}}_G \sim N\left(\mathbf{0}, (1+3\nu)\,\sigma\cdot\mathbf{I}\right)$, using codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$, satisfying

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{f}}_G\right]\right] = \mathbb{P}\left[\left\{\exists k \neq i \,:\, \left\|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{f}}_G\right\| \leq \left\|\tilde{\mathbf{f}}_G\right\|\right\}\right] \leq \overline{\epsilon}.$$

We begin by showing that the codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ satisfy constraints (25). To show that, consider the probability of incorrectly decoding $\mathbf{x}_i$ as $\mathbf{x}_k$ on this channel. We have

$$\begin{aligned}
\mathbb{P}\left[\mathbf{x}_i \text{ decoded as } \mathbf{x}_k\right] &\geq \mathbb{P}\left[\left\|\mathbf{x}_i + \tilde{\mathbf{f}}_G - \mathbf{x}_k\right\| \leq \left\|\tilde{\mathbf{f}}_{\mathbf{G}}\right\|\right] \\
&= \mathbb{P}\left[2\left\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \tilde{\mathbf{f}}_G\right\rangle \geq \left\|\mathbf{x}_i - \mathbf{x}_k\right\|^2\right] \\
&= \mathbb{P}\left[\frac{\left\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \tilde{\mathbf{f}}_G\right\rangle}{(1+3\nu)\,\sigma\left\|\mathbf{x}_i - \mathbf{x}_k\right\|} \geq \frac{\left\|\mathbf{x}_i - \mathbf{x}_k\right\|}{2\,(1+3\nu)\,\sigma}\right] \\
&= 1 - \Phi\left(\frac{\left\|\mathbf{x}_i - \mathbf{x}_k\right\|}{2\,(1+3\nu)\,\sigma}\right).
\end{aligned}$$

Since we know that $\mathbb{P}\left[\mathbf{x}_i \text{ decoded as } \mathbf{x}_k\right] \leq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{f}}_G\right]\right] \leq \overline{\epsilon}$, we have

$$\begin{aligned}
&\overline{\epsilon} \geq 1 - \Phi\left(\frac{\left\|\mathbf{x}_i - \mathbf{x}_k\right\|}{2\,(1+3\nu)\,\sigma}\right) \\
&\implies \left\|\mathbf{x}_i - \mathbf{x}_k\right\| \geq 2\,(1+3\nu)\,\sigma\Phi^{-1}\left(1-\overline{\epsilon}\right) = 2\,(1+3\nu)\,\zeta\sqrt{n}, \qquad (33)
\end{aligned}$$

which implies that the codewords satify the constraints (25).

We next show that the codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ satisfy Eqs. (23) and (24). In particular, we need to show that the number of $\mathbf{z}_t$ that lead to a decoding error is bounded by ϵ, that is, we have to show that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\left\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\right\| \leq \left\|\mathbf{z}_t\right\|\right\} \leq \epsilon,$$

where $\mathbf{1}\left\{\left\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\right\| \leq \left\|\mathbf{z}_t\right\|\right\}$ denotes that a decoding error occured at noise vector $\mathbf{z}_t$.

Note that $\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\left\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\right\| \leq \left\|\mathbf{z}_t\right\|\right\}$ is the error probability when the noise is distributed uniformly in the set $\{\mathbf{z}_t\}_{t=1}^{T}$. Noting this, we choose to analyze the error probability of this code under a channel in which the noise is uniformly distributed in $\mathcal{S}_n(\sigma_1\sqrt{n})$, where $\sigma_1 = (1+2\nu)\,\sigma$. Let us call it the noise $\tilde{\mathbf{z}}_U$. We will next show that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\left\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\right\| \leq \left\|\mathbf{z}_t\right\|\right\} \leq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_U\right]\right],$$

which helps us bound the quantity $\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\left\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\right\| \leq \left\|\mathbf{z}_t\right\|\right\}$.

In order to show this, consider any $\mathbf{z}_t$, and let $\mathbf{u}_t = \sigma_1 \mathbf{z}_t / (1+\nu)\sigma$. We have $\|\mathbf{u}_t\| = \sigma_1\sqrt{n}$. In the first step, we show that

$$\text{if } \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|, \text{ then } \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{s}\| \leq \|\mathbf{s}\| \ \forall \mathbf{s} \in \mathcal{V}(\mathbf{u}_t).$$

Let $\mathbf{s} \in \mathcal{S}_n(\sigma_1\sqrt{n})$ such that $\mathbf{s}$ is in the Voronoi region $\mathcal{V}(\mathbf{u}_t)$ of $\mathbf{u}_t$, that is, $\mathbf{s} \in \mathcal{V}(\mathbf{u}_t)$. Applying Proposition 4 to $\mathcal{A} = \{\mathbf{u}_1, \ldots, \mathbf{u}_T\}$, $N = T$, $\Lambda = \sigma_1$, we obtain

$$\|\mathbf{s} - \mathbf{u}_t\| \leq \theta' \sqrt{n}, \tag{34}$$

where

$$\begin{aligned}
\theta' &= \frac{\sigma_1}{T^{1/n}} = \frac{\sigma_1\sqrt{n}}{\gamma_\epsilon} \cdot \frac{\gamma_\epsilon}{\sqrt{n}T^{1/n}} = \frac{\sigma_1\sqrt{n}}{\gamma_\epsilon} \cdot \zeta \cdot \frac{\nu}{1+\nu} \\
&\leq \frac{\sigma_1\sqrt{n}}{\gamma_\epsilon} \cdot \frac{1}{1+2\nu} \cdot \frac{1}{2\sqrt{n}} \cdot \|\mathbf{x}_i - \mathbf{x}_k\| \cdot \frac{\nu}{1+\nu} \quad \text{(from (33))} \\
&= \frac{(1+2\nu)\,\sigma\sqrt{n}}{\gamma_\epsilon} \cdot \frac{1}{1+2\nu} \cdot \frac{\nu}{1+\nu} \cdot \frac{1}{2\sqrt{n}} \cdot \|\mathbf{x}_i - \mathbf{x}_k\| \leq \frac{\nu}{1+\nu} \cdot \frac{1}{2\sqrt{n}} \cdot \|\mathbf{x}_i - \mathbf{x}_k\|.
\end{aligned} \tag{35}$$

Finally, if $\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|$, then

$$\begin{aligned}
2\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \mathbf{s}\rangle &= 2\langle \mathbf{x}_k - \mathbf{x}_i, (\mathbf{s} - \mathbf{u}_t) + \mathbf{u}_t\rangle \\
&\geq 2\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \mathbf{u}_t\rangle - 2 \cdot \|\mathbf{x}_i - \mathbf{x}_k\| \cdot \|\mathbf{s} - \mathbf{u}_t\| \quad \text{(Cauchy-Schwartz)} \\
&\geq \frac{1+2\nu}{1+\nu}\|\mathbf{x}_i - \mathbf{x}_k\|^2 - 2 \cdot \|\mathbf{x}_i - \mathbf{x}_k\| \cdot \theta'\sqrt{n} \quad \text{(from (34))} \\
&= \|\mathbf{x}_i - \mathbf{x}_k\| \cdot \left\{ \frac{1+2\nu}{1+\nu}\|\mathbf{x}_i - \mathbf{x}_k\| - 2\theta'\sqrt{n} \right\} \\
&\geq \|\mathbf{x}_i - \mathbf{x}_k\| \cdot \left\{ \frac{1+2\nu}{1+\nu}\|\mathbf{x}_i - \mathbf{x}_k\| - \frac{\nu}{1+\nu} \cdot \|\mathbf{x}_i - \mathbf{x}_k\| \right\} \quad \text{(from (35))} \\
&= \|\mathbf{x}_i - \mathbf{x}_k\|^2,
\end{aligned}$$

which is equivalent to $\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{s}\| \leq \|\mathbf{s}\| \ \forall \mathbf{s} \in \mathcal{V}(\mathbf{u}_t)$. Therefore,

$$\text{if } \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|, \text{ then } \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{s}\| \leq \|\mathbf{s}\| \ \forall \mathbf{s} \in \mathcal{V}(\mathbf{u}_t). \tag{36}$$

Furthermore,

$$\mathbb{P}[\mathscr{E}_i[\tilde{\boldsymbol{z}}_U]] = \sum_{t=1}^{T} \mathbb{P}\left[\mathscr{E}_i[\tilde{\boldsymbol{z}}_U] \,\middle|\, \tilde{\boldsymbol{z}}_U \in \mathcal{V}(\mathbf{u}_t)\right] \cdot \mathbb{P}[\tilde{\boldsymbol{z}}_U \in \mathcal{V}(\mathbf{u}_t)]. \tag{37}$$

Since the set of vectors $\{\mathbf{z}_t\}$ form a Voronoi tessellation, the vectors $\{\mathbf{u}_t\}$ also form a Voronoi tessellation on $\mathcal{S}_n(\sigma_1\sqrt{n})$. Therefore, the Voronoi regions of the points $\mathbf{u}_t$ are identical with the same area. Consequently,

$$\mathbb{P}[\tilde{\boldsymbol{z}}_U \in \mathcal{V}(\mathbf{u}_t)] = \frac{1}{T}, \ \forall t = 1, \ldots, T. \tag{38}$$

Moreover, from (36), we have that there exists a k such that

$$\mathbb{P}\left[\mathscr{E}_i[\tilde{\boldsymbol{z}}_U] \,\middle|\, \tilde{\boldsymbol{z}}_U \in \mathcal{V}(\mathbf{u}_t)\right] \geq \mathbf{1}\{\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|\}. \tag{39}$$

Substituting (38) and (39) in (37), we have

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_U\right]\right] \geq \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|\right\}. \tag{40}$$

From Proposition 7 we have,

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_U\right]\right] \leq \frac{\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_G\right]\right]}{1-\delta\left(\nu,n\right)} \leq \frac{\overline{\epsilon}}{1-\delta\left(\nu,n\right)} = \epsilon. \tag{41}$$

Therefore, from (40), we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| \leq \|\mathbf{z}_t\|\right\} \leq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_U\right]\right] \leq \epsilon,$$

which implies that the codewords satisfy constraints (24). Therefore, we have shown that the codewords satisfy all the constraints of the optimization problem, implying that $r^* = 1$. This proves the contra-positive statement of Part(b) thus proving it. □

Recall that the asymptotic capacity region $\mathcal{R}\left[P,\sigma\right]$ is defined as

$$\mathcal{R}\left[P,\sigma\right] = \lim_{n\to\infty}\mathcal{R}_n\left[P,\sigma,\epsilon_n\right], \text{ where } \epsilon_n \to 0, \text{ as } n \to \infty.$$

Theorem 1 provides bounds on the channel capacity and a code that matches the lower bound for finite n and ϵ, as well as explicit bounds for the error probabilities given n and ϵ. In the limit $n \to \infty, \epsilon \to 0$ the lower and upper bounds become tight.

From a computational point of view, we need to solve large scale NP-hard problems to find lower and upper bounds for the channel capacity. However, we report computational evidence in Section 7 that suggests we can solve problems with $n = 140$.

4 Channels with Additive Non-Gaussian Noise

In this section, we explore how the nature of the optimization problem we solve to compute the capacity region and to find a matching code depends on the specific probabilistic assumptions we make on the noise of the channel. In previous sections, we have seen that if the noise is Gaussian, the underlying optimization problem becomes a rank minimization problem with semidefinite constraints. In Section 6.1, we show that when the noise is exponentially distributed, then the underlying capacity computation problem for single-user channels is a mixed binary linear optimization problem. In Section 6.2, we explore the cases of uniform, binary symmetric noise and the case where we make no specific probabilistic assumption on the noise, but assume that the noise sequences satisfy certain limit laws.

4.1 Single-User Channel with Additive Exponentially Distributed Noise

We consider both the single-user channel when the noise is exponentially distributed. We begin by considering a single-user channel, where we intend to construct a code $\mathcal{C}$ consisting of a codebook $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ of size M. We next present the maximum likelihood decoder for the additive exponential noise channel.

Proposition 9. *Consider a code $\mathcal{C}$ for a single-user additive exponential noise channel with codebook $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$. The maximum likelihood decoder is given by $g_0^E(\mathbf{y}) = \arg\min_{i \in \mathcal{B}(\mathbf{y})} \sum_{j=1}^n (y_j - x_{ij})$, where $\mathcal{B}(\mathbf{y}) = \left\{ i \in \mathcal{B} \middle| y_j \geq x_{ij},\ \forall j = 1, \ldots, n \right\}$.*

We next present the optimization problems that we use to characterize the capacity regions of the additive exponential noise channels. In this direction, we observe from Corollary 1(b) that the typical set for an exponential distribution with parameter λ is given by

$$\mathcal{U}_\epsilon^E = \left\{ (z_1, \ldots, z_n) \middle| \frac{n}{\lambda} - \sqrt{n}\frac{\Gamma_\epsilon^E}{\lambda} \leq \sum_{j=1}^n z_j \leq \frac{n}{\lambda} + \sqrt{n}\frac{\Gamma_\epsilon^E}{\lambda} \right\},$$

where Γ_ϵ^E is chosen such that $\mathbb{P}\left[\tilde{\mathbf{z}} \in \mathcal{U}_\epsilon^E\right] = 1 - \epsilon$, when each component of $\tilde{\mathbf{z}}$ is distributed exponentially with parameter λ.

Motivated by the decoder, we next present a RO problem (42) that allows us to characterize the capacity region of a single-user additive exponential noise channel. In particular, given inputs $n, R, \lambda, P, \epsilon, \nu$, we calculate the "derivative" quantities γ_ϵ^E, T^E and M_0^E as follows:

1. The parameter γ_ϵ^E, which we choose so that $\mathbb{P}\left[\sum_{i=1}^n \tilde{z}_i \leq \gamma_\epsilon\right] \geq 1 - \epsilon$, where $\tilde{z}_i \sim$ exponential(λ).

2. The parameter T^E given by

$$T = \left(\frac{1 + 2\nu}{\zeta^E \nu} \cdot \frac{\gamma_\epsilon^E}{\sqrt{n}} \right)^n, \text{ with } \zeta^E = \frac{1}{\lambda\sqrt{n}} \cdot \Psi^{-1}(1 - \epsilon),$$

where $\Psi(\cdot)$ is the cdf of the exponential distribution;

3. The parameter $M_0^E = (1 + 2\nu) \cdot \gamma_\epsilon^E$. In addition, we generate a Voronoi tessellation $\left\{\mathbf{z}_1^E, \mathbf{z}_2^E, \ldots, \mathbf{z}_T^E\right\}$ of the simplex

$$\mathcal{P}_\epsilon = \left\{ (z_1, \ldots, z_n) \middle| \sum_{j=1}^n z_j = \frac{n}{\lambda} + \sqrt{n}\frac{\Gamma_\epsilon^E}{\lambda} \right\}.$$

Let $\mathcal{M} = \{1, \ldots, 2^{nR}\}$ and $\mathcal{T} = \{1, \ldots, T\}$. We next use the decision variables $\mathbf{x}_i$, $i \in \mathcal{M}$ and v_{it}, $i \in \mathcal{M},\ t \in \mathcal{T}$, where

(a) The variables $\mathbf{x}_i$ represent the codewords.

(b) The variables v_{it} represent binary decision variables that are chosen in a way to constrain the probability of error. When $v_{it} = 1$, the set of decoding constraints in (42) are satisfied for codeword $\mathbf{x}_i$ with noise vector $\mathbf{z}_t^E$. We construct the following mixed binary linear optimization problem:

$$\begin{aligned}
\max \quad & \sum_{i,k,t} v_{ikt} && && (42)\\
\text{s.t.} \quad & \sum_{j=1}^n x_{ij} \leq nP, && \forall i = 1, \ldots, 2^{nR},\\
& \sum_{j=1}^n x_{ij} + (2 - v_{it} - v_{ikt}) M_0 \geq \sum_{j=1}^n x_{kj}, && \forall t, \forall i, k \neq i,\\
& x_{ij} + z_{tj}^E \geq x_{kj} - M_0(1 - v_{ikt}), && \forall i, k, j, t,\\
& \sum_{t=1}^T v_{it} \geq (1 - \epsilon) T, && \forall i,\\
& v_{it}, v_{ikt} \in \{0, 1\}, && \forall i, k, t,
\end{aligned}$$

We next present Algorithm 2, that computes the capacity region $\mathcal{R}_n^E[P,\lambda,\epsilon]$ of this channel.

Algorithm 2. ***Capacity Computation and Optimal Coding for the Single User additive exponential noise channel.***
Input: *R, P, λ, n, ν, ϵ.*
Output: *Codewords$\{\mathbf{x}_i\}_{i\in\mathcal{M}}$, and auxiliary binary variables $\{v_{it}, v_{ikt}\}$.*
Algorithm:

1. *Solve the mixed binary linear optimization problem (42) to compute the codewords$\{\mathbf{x}_i\}_{i\in\mathcal{M}}$, and auxiliary binary variables $\{v_{it}, v_{ikt}\}$.*
2. *If the problem is feasible, then $R \in \mathcal{R}_n^E[P,\lambda,2\epsilon]$, that is, R can be achieved on an additive exponential noise channel using the codewords $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ and the decoding function $g_0^E(\cdot)$, with a decoding error probability of at most ϵ.*
3. *If the problem is infeasible, then $R \notin \mathcal{R}_n^E\left[P, \dfrac{\lambda}{1+2\nu}, 3\epsilon\right]$.*

Theorem 2. ***(Capacity Region in a Single-User additive exponential noise channel)***

(a) *If problem (42) is feasible, then $R \in \mathcal{R}_n^E[P,\lambda,2\epsilon]$, that is, R is achievable using the codebook $\mathcal{B} = \{\mathbf{x}_i\}_{i=1}^{2^{nR}}$, and the maximum likelihood decoder (Proposition 9), achieving an average decoding error probability of 2ϵ.*

(b) *If problem (42) is infeasible, then $R \notin \mathcal{R}_n^E\left[P, (1+2\nu)^{-1}\lambda, 3\epsilon\right]$.*

The proof of Theorem 2 is similar to the case of Gaussian noise (Theorem 1) and is omitted.

5 Computational Results

In this section, we discuss the computational complexity of our approach and present computational results to illustrate the effectiveness of the RO approach for the single-user Gaussian and additive exponential noise channels.

5.1 Computational Complexity

We begin by discussing the computational complexity of the algorithms we presented. In Sections 3, we have shown that the optimization problem for Gaussian channels is a rank minimization problem with semidefinite constraints. On the other hand, in Section 4.1, we have shown that for additive exponential noise channels, the optimization problem is a mixed binary linear optimization problem. We next present the size of these optimization problems as a function of the key problem parameters:

1. *Additive Gaussian noise channel*: Recall that the key parameters that characterize the channel coding problem (22–26) are the code-length n, noise standard deviation σ, signal power P, rate R, error probability ϵ and approximation parameter ν. Given these parameters, the number of variables is given by $\mathcal{O}\left(2^{nR} + \left(\frac{(1+\nu)}{\sigma\nu}\cdot\Phi^{-1}(1-\epsilon)\right)^n\right)$ which is exponential in n, and depends on the accuracy ν required, error probability ϵ and the rate R.

2. *Additive Exponential noise channel*: The key parameters that characterize the channel coding problem (42) are the code-length n, noise rate λ, signal power P, rate R, error probability ϵ and approximation parameter ν. Given these parameters, the number of variables is given

by $\mathcal{O}\left(2^{nR} + \left(\frac{(1+2\nu)\lambda}{\nu} \cdot \Psi^{-1}(1-\epsilon)\right)^n\right)$, where $\Psi(\cdot)$ is the exponential cdf. Note that this is exponential in n, and depends on the accuracy ν required, error probability ϵ and the rate R.

Nature of the Optimization Problem and Structure of the Typical Set

Furthermore, the nature of the optimization problem depends on the type of the typical set. The key observation is that when the typical set is a polyhedron, the underlying optimization problem is a binary linear optimization problem (either mixed or not).

In Corollary 2, we characterized the typical sets for various distributions. In particular, we showed that the typical sets for the uniform and binary distributions are polyhedra. Moreover, if we do not know the specific distribution of noise in a channel, we can potentially use limit laws that random sequences satisfy, to model the primitives of the noise. Specifically, we may assume the following uncertainty sets for noise sequences $\mathbf{X} = \{X_1, \ldots, X_n\}$:

1. *The Central Limit Theorem (CLT)* : The Central Limit Theorem states that the normalized sum of random variables $(X_1 + \ldots + X_n - n\mu)/(\sigma \cdot \sqrt{n})$ is asymptotically standard normal. This allows us to construct an uncertainty set that makes use of the properties of normal random variables. In particular, we can construct an uncertainty set $\mathbb{U}^{\text{CLT}}$ as follows

$$\mathcal{U}^{\text{CLT}} = \left\{ (X_1, X_2, \ldots, X_n) \,\middle|\, -\Gamma \leq \left(\sum_{i=1}^{n} X_i - n\mu\right) / \sigma \cdot \sqrt{n} \leq \Gamma \right\}, \tag{43}$$

   where Γ is chosen using the properties of the normal distribution, as we discussed earlier in this section.

2. *Stable Laws* : These limit laws express the fact that a sum of many independent random variables will tend to be distributed according to one of a small set of "attractor" (i.e. stable) distributions. When the variance of the variables is finite, the "attractor" distribution is the normal distribution. In particular, these stable laws allow us to construct uncertainty sets for heavy-tailed distributions. This allows us to consider the uncertainty set

$$\mathcal{U}^{\text{HT}} = \left\{ (X_1, X_2, \ldots, X_n) \,\middle|\, -\Gamma \leq \left(\sum_{i=1}^{n} X_i - n\mu\right) / n^{1/\alpha} \leq \Gamma \right\}, \tag{44}$$

where Γ can be chosen based on the distributional properties of the heavy tailed random variables X_i. Using the RO approach used in Sections 3, and 4.1, we present a summary of the nature of the optimization problems in Table 1. From Table 1, we see that for many cases the underlying optimization problem is a mixed binary linear optimization problem. Due to advances in optimization theory and processing speeds in the last three decades, large scale mixed binary linear optimization problems are solved routinely by commercial solvers, where as large scale semidefinite optimization problems present computational challenges. Given that, in reality, we have a choice in modeling noise, it is reasonable in our opinion, to model noise so that the underlying typical set are polyhedra. In particular, when we model noise using the limit laws outlined in Eqs. (43) and (44), the resulting optimization problems become mixed binary linear optimization problems.

Noise	Typical Set	Optimization Problem
Gaussian (independent)	Ball in (8)	Rank minimization with semidefinite constraints
Gaussian (correlated)	Ellipsoid in (9)	Rank minimization with semidefinite constraints
exponential	Polyhedron in (10)	Mixed binary linear optimization problem
uniform	Polyhedron in (11)	Mixed binary linear optimization problem
Binary symmetric noise	Polyhedron in (12)	Binary optimization problem
CLT Uncertainty Set	Polyhedron in (43)	Mixed binary linear optimization problem
Stable Law Uncertainty Set	Polyhderon in (44)	Mixed binary linear optimization problem

Table 1: Dependence of the nature of the optimization problem with noise models.

5.2 Implementation

For single-user Gaussian channels we presented Algorithm 1 for encoding that involve the solution of a rank minimization problem subject to semidefinite constraints:

$$\begin{array}{ll} \min & \operatorname{rank}(\mathbf{X}) \\ \text{s.t.} & \tilde{\mathbf{A}} \bullet \mathbf{X} \leq \mathbf{0}, \\ & \mathbf{X} \succeq \mathbf{0}, \end{array} \tag{45}$$

We use the following iterative algorithm developed by Fazell et al. [2003], to solve Problem (45).

Algorithm 3. ***Solving Rank Minimization with Semidefinite Constraints***
Input: $\tilde{\mathbf{A}}, \delta_0, K$.
Output: *A matrix* $\mathbf{X}^K$*, solution to Problem (45).*
Algorithm:

1. *Solve the convex optimization problem*

$$\begin{array}{ll} \min & Tr(\mathbf{X}) \\ s.t. & \tilde{\mathbf{A}} \bullet \mathbf{X} \leq \mathbf{0}, \\ & \mathbf{X} \succeq \mathbf{0}, \end{array}$$

and let $\mathbf{X}^0$ *denote the optimal solution.*

2. *For each iteration* $k = 1, \ldots, K$*, set* $\delta = \delta_0/k$ *and solve the optimization problem*

$$\begin{array}{ll} \min & Tr\left(\left(\mathbf{X}^{k-1} + \delta I\right)^{-1} \mathbf{X}\right) \\ s.t. & \tilde{\mathbf{A}} \bullet \mathbf{X} \leq \mathbf{0}, \\ & \mathbf{X} \succeq \mathbf{0}. \end{array}$$

The key connection of Problem (45) and Algorithm 3 is the formula:

$$\operatorname{rank}(\mathbf{X}) = n - \lim_{\delta \to 0} \frac{\log\det(\mathbf{X} + \delta \mathbf{I})}{\log \delta},$$

and thus in order to solve Problem (45) we aim to minimize $\log\det(\mathbf{X} + \delta\mathbf{I})$ with successively decreasing values of δ. Algorithm 3 can be interpreted as a steepest descent algorithm on $\log\det(\mathbf{X} + \delta\mathbf{I})$ with successively decreasing values of δ. Fazell et al. [2003] showed that Algorithm 3 is guaranteed to converge to a local minimum of $\log\det(\mathbf{X} + \delta\mathbf{I})$. The empirical behavior of Algorithm 3 depends on the choices of the parameters δ_0 and K. Yu and Lau [2011], Wang and Sha [2011] report that Algorithm 3 finds the minimum rank successfully in signal processing applications.

Remark

Note that when we find $r^* = 1$, we get a code and therefore, we establish a valid lower bound. On the other hand, a proof of infeasibility of $r^* = 1$ is difficult. But recall that, since we formulate the encoding problem (29) as a binary semidefinite problem, we can in principle, produce a certificate of infeasibility by enumerating all possible binaries and then optimize over the remaining semidefinite constraints. Clearly this is not a practical method but it does illustrate that, in principle, our approach provides valid certificates but the process can take potentially exponential time. In our implementation, to check $r^* \geq 2$, we try 50,000 restarts and declare the unachievability if we fail in each of these restarts. In this sense, the upper bounds we report are approximate, while the lower bounds are exact.

5.3 The Single-User Gaussian Channel

For a single-user Gaussian channel and use Algorithm 1 to compute the capacity region. In order to construct the capacity region, we choose values of $n \leq 140, \epsilon = 0.001, \nu = 0.05$ and for different values of R, we check whether the code construction problem is feasible ($r^* = 1$ or $r^* \geq 2$). As discussed before, the size of the problem depends on the value of the rate R. For example, for $R = 0.2$ the resulting semidefinite optimization problems for $n = 140$ involves around 3 billion variables. In order to solve this problem, we divide this problem by dividing the feasible region into 5000 equal regions. In each region, we seek to find $2^{nR}/5000$ codewords. Let $\mathcal{A}_k$ be the set of codewords in the k^{th} region, where each of the codeword in this region satisfies the linear constraint

$$cos\left((k-1)\pi/5000\right) \leq \mathbf{x}_{i1} \leq cos\left(k\pi/5000\right), \quad \forall \mathbf{x}_i \in \mathcal{A}_k, \tag{46}$$

where $cos(\cdot)$ is the cosine function. The main problem now reduces to 5000 subproblems each of which has 600 thousand variables. Each of the subproblem has the extra constraint (46). We use the open source implementation of the SDPARA algorithm (Yamashita et al. [2003]) which allows parallelization. This algorithm took 18 hours on a multi-core linux machine with 48GB RAM and 8 processors. For the maximum value of $R = 0.35$ that we computed, the total number of variables exceeds 50 trillion variables. In order to solve this problem, we divided the problem into 5 million equal regions as in Eq. (46). Each subproblem involved 2 million variables. This entire exercise tool more than 700 hrs on a 32-core linux machine with 168 GB RAM using the SDPARA algorithm.

In Figure 1, and for specific values of n we provide the lower and upper bounds from the RO approach, that is an interval $[\underline{R}, \overline{R}]$ such that $\underline{R}$ is achievable and we construct the corresponding code, while $\overline{R}$ is not achievable by any code. For comparison, we present comparable lower and upper bounds using the methodology in Polyanskiy et al. [2010]. We also record the asymptotic

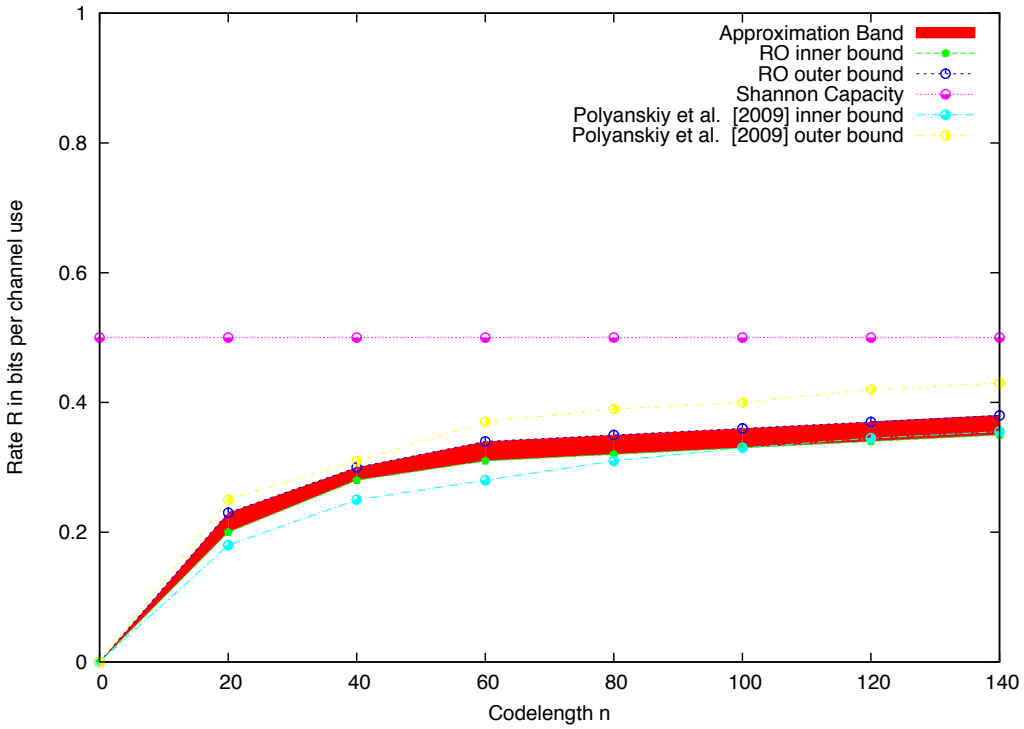

Figure 1: Comparison of the lower and upper bounds provided by the RO approach and by Polyanskiy et al. [2010] as a function of the code length n in a single-user Gaussian channel.

Shannon capacity. We observe that the upper bound of the RO approach is sharper than the upper bound in Polyanskiy et al. [2010], while the lower bounds are comparable. Furthermore, for $n = 140$, the asymptotic Shannon capacity is still quite far from the lower and upper bounds we achieve.

5.4 The Single-User additive exponential noise channel

In this section, we aim to examine whether we are able to solve problems with larger code length n if the typical sets are polyhedra as opposed to ellipsoids. We selected a single-user additive exponential noise channel and applied Algorithm 2. We were able to find lower and upper bounds on channel capacity for values of $n = 300, \epsilon = 0.001, \nu = 0.05$, compared to $n = 140$ for the Gaussian case. The mixed binary linear optimization problems we solved involved $1,430,000$ variables. We used CPLEX 11.1 on a computer with 150G RAM, 8 processors running on LINUX. To check whether a given R is achievable took 18 hours for $n = 300$. The ability to solve larger problems in this case is due to the fact that the state of the art in computational linear integer optimization is more advanced than rank minimization with semidefinite constraints, which is the key computational problem for Gaussian channels.

In Figure 2, we report the lower and upper bounds from the RO approach as a function of n. For comparison, we also record the asymptotic Shannon capacity in this case.

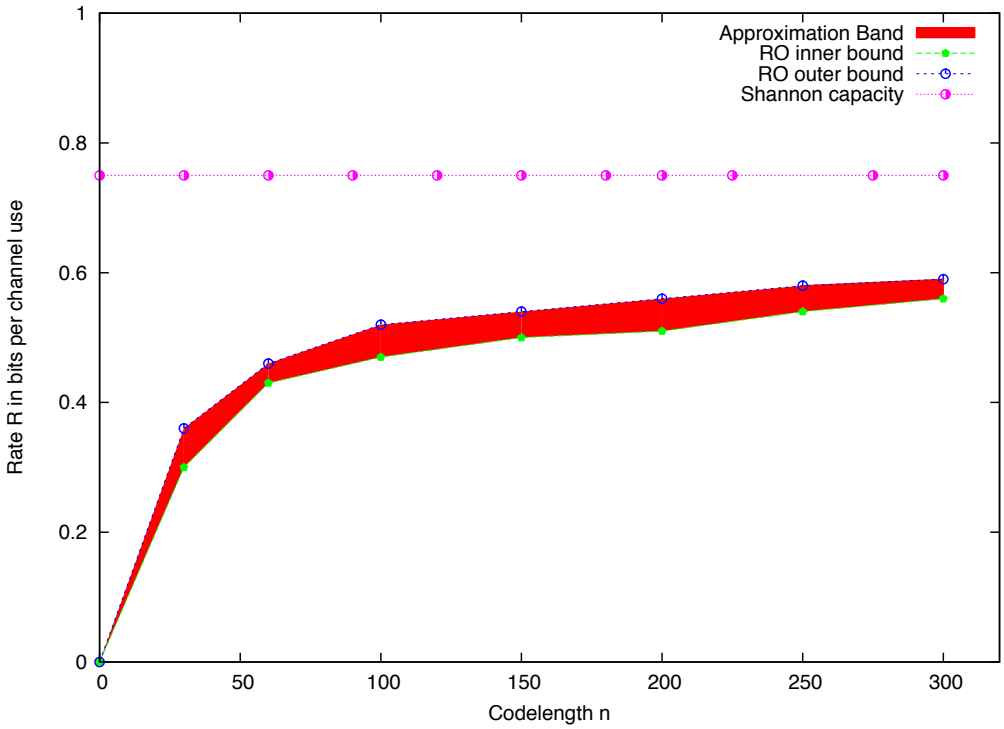

Figure 2: Lower and upper bounds provided by the RO approach as a function of the code length n in a single-user exponential channel.

6 Conclusions

In this paper, we proposed a RO approach to formulate and algorithmically solve problems in network information theory. As summarized in Table 1, the nature of the optimization problem ranges from binary/mixed binary linear optimization problems to rank minimization problems with semidefinite constraints. We have been able to solve problems with $n = 140$ for single-user Gaussian channels, and $n = 300$ for single-user exponential channels. The sizes of the problems we can solve for non-Gaussian channels are in fact larger as the state of the art in computational linear integer optimization is more advanced than rank minimization with semidefinite constraints, which is the key computational problem for Gaussian channels. Given that we have a choice in modeling noise, it is reasonable in our opinion, to model noise so that the underlying typical set are polyhedra. Furthermore, as optimization algorithms and computing infrastructure improve, we anticipate we will be able to increase the size of the problems we are to able to tackle potentially significantly.

Acknowledgements

We thank the Associate editor Professor Guo and the referees for insightful comments that improved the paper significantly.

References

C. Bandi and D. Bertsimas. Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134(1):23–70, 2012.

C. Bandi and D. Bertsimas. Optimal design for multi-item auctions: A robust optimization approach. *Mathematics of Operations Research*, 39(4):1012–1038, 2014a.

C. Bandi and D. Bertsimas. Robust option pricing. *European Journal of Operations Research*, 239(3):842–853, 2014b.

C. Bandi and D. Bertsimas. Channel coding via robust optimization, part 2: The multi-channel case. *Submitted for publication*, 2015.

C. Bandi, D. Bertsimas, and N. Youssef. Robust queueing theory. *Operations Research, to appear*, 2015.

A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

A. Ben-Tal and A. Nemirovski. Robust solutions to uncertain programs. *Operations Research Letters*, 25:1–13, 1999.

A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

A. Ben-Tal, L. El-Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98:49–71, 2003.

D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Optimization* . Athena Scientific and Dynamic Ideas, Belmont, Feb. 1997. ISBN 1886529191.

D. Bertsimas and R. Weismantel. *Optimization Over Integers*. Dynamic Ideas, Belmont, 2005. ISBN 0975914626. URL `http://www.worldcat.org/isbn/0975914626`.

D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53:464–501, 2011.

T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, New York, NY, USA, 2006.

I. Cplex. 11.0, 2014.

G. B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis of Production and Allocation*, pages 339–347, 1947.

Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):636–676, 1999.

Q. Du, M. Emelianenko, and L. Ju. Convergence of the Lloyd algorithm for computing centroidal voronoi tessellations. *SIAM Journal on Numerical Analysis*, 44(1):102–119, 2006.

L. El-Ghaoui and H. Lebret. Robust solutions to least-square problems to uncertain data matrices. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.

L. El-Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9:33–52, 1998.

M. Fazell, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. *Proceedings American Control Conference*, 3: 2156–2162, 2003.

R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6): 2325–29, 1998.

Gurobi. Gurobi 4.0.2. software, Dec. 2010.

Y. Liu, W. Wang, B. Lévy, F. Sun, D.-M. Yan, L. Lu, and C. Yang. On centroidal voronoi tessellation-energy smoothness and fast computation. *ACM Transactions on Graphics (ToG)*, 28(4):101, 2009.

S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2):129–137, 1982.

Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *Information Theory, IEEE Transactions on*, 56(5):2307–2359, 2010.

M. Sabin and R. M. Gray. Global convergence and empirical consistency of the generalized Lloyd algorithm. *IEEE Transactions on Information Theory*, 32(2):148–155, 1986.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27: 379–423, 1948.

S. Verdú. Fifty years of shannon theory. *IEEE Transactions on information theory*, 44(6):2057–2078, 1998.

S. Verdú and S. W. McLaughlin, editors. *Information Theory: 50 Years of Discovery*. IEEE Press, Piscataway, NJ, USA, 2000. ISBN 0-7803-5363-3.

M. Wang and F. Sha. Information theoretical clustering via semidefinite programming. *AISTATS*, pages 761–769, 2011.

A. Wyner. Random packing and coverings of the unit n-sphere. *Bell System Technical Journal*, 46(9):2111–2118, 1967.

M. Yamashita, K. Fujisawa, and M. Kojima. Sdpara: Semidefinite programming algorithm parallel version. *Parallel Computing*, 29(8):1053–1067, 2003.

H. Yu and V. K. Lau. Rank-constrained schur-convex optimization with multiple trace/log-det constraints. *IEEE Transactions on Signal Processing*, 59(1):304–314, 2011.

Appendix A. Proofs of Auxilliary Results

Proof of Proposition 6.
(a) We first show that

$$\text{if } \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}\| \leq \|\tilde{\mathbf{z}}\|, \text{ then } \|\mathbf{x}_i - \mathbf{x}_k + \alpha\tilde{\mathbf{z}}\| \leq \alpha \|\tilde{\mathbf{z}}\|, \forall \alpha \geq 1. \tag{47}$$

If $\|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}\| \leq \|\tilde{\mathbf{z}}\|$, then we have

$$\begin{aligned}\|\mathbf{x}_i - \mathbf{x}_k + \alpha\tilde{\mathbf{z}}\| &= \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}} + (\alpha - 1)\tilde{\mathbf{z}}\| \\ &\leq \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{z}}\| + \|(\alpha - 1)\tilde{\mathbf{z}}\| \quad \text{(triangle inequality)} \\ &\leq \|\tilde{\mathbf{z}}\| + (\alpha - 1)\|\tilde{\mathbf{z}}\| = \alpha\|\tilde{\mathbf{z}}\|.\end{aligned}$$

We next consider a sample path ω_B such that $\exists \tilde{k} \neq i : \|\mathbf{x}_i - \mathbf{x}_{\tilde{k}} + \tilde{\boldsymbol{u}}_B(\omega_B)\| \leq r_B$. Then, consider a sample path ω_A given by

$$\tilde{\boldsymbol{u}}_A(\omega_A) = \frac{r_A}{r_B} \cdot \tilde{\boldsymbol{u}}_B(\omega_B).$$

Applying (47) with $\alpha = r_A/r_B \geq 1$, we have $\|\mathbf{x}_i - \mathbf{x}_{\tilde{k}} + \alpha\tilde{\boldsymbol{u}}_B(\omega_B)\| \leq \alpha r_B$, leading to

$$\|\mathbf{x}_i - \mathbf{x}_{\tilde{k}} + \tilde{\boldsymbol{u}}_A(\omega_A)\| \leq r_A.$$

In other words, we have

$$\mathbf{1}\{\mathscr{E}_i[\tilde{\boldsymbol{u}}_A(\omega_A)]\} \geq \mathbf{1}\{\mathscr{E}_i[\tilde{\boldsymbol{u}}_B(\omega_B)]\}. \tag{48}$$

Noting that, if $\tilde{\boldsymbol{u}}_B(\omega_B)$ is uniformly distributed in $\mathcal{S}_n(r_B)$, then $\tilde{\boldsymbol{u}}_A(\omega_A)$ is also uniformly distributed in $\mathcal{S}_n(r_A)$, by taking expectations in (48), we have

$$\mathbb{P}\Big[\exists k \neq i : \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_A\| \leq r_A\Big] \geq \mathbb{P}\Big[\exists k \neq i : \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\boldsymbol{u}}_B\| \leq r_B\Big].$$

(b) Let $\mathbf{s} \in \mathcal{S}_n(1)$, and let $f_i, g_i : \mathcal{S}_n(1) \to \{0,1\}$ be defined as

$$f_i(\mathbf{s}) = \begin{cases} 1, & \text{if } \exists d \leq r_A : \|\mathbf{x}_i - \mathbf{x}_k + d\mathbf{s}\| \leq d, \text{ for some } k \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

$$g_i(\mathbf{s}) = \begin{cases} 1, & \text{if } \|\mathbf{x}_i - \mathbf{x}_k + r_A\mathbf{s}\| \leq r_A, \text{ for some } k \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

Applying (47) with $\tilde{\mathbf{z}} = d \cdot \mathbf{s}$ and $\alpha = r_A/d \geq 1$, we have that if $f_i(\mathbf{s}) = 1$, then $g_i(\mathbf{s}) = 1$ leading to

$$\mathbb{P}_{\mathbf{s}\sim\tilde{\mathbf{s}}(1)}[g_i(\mathbf{s}) = 1] \geq \mathbb{P}_{\mathbf{s}\sim\tilde{\mathbf{s}}(1)}[f_i(\mathbf{s}) = 1]. \tag{49}$$

We next consider the event $\Big\{\mathscr{E}_i[d\mathbf{s}] \Big| d = r_A\Big\}$. We have

$$\Big\{\mathscr{E}_i[d\mathbf{s}] \Big| d = r_A\Big\} \iff g_i(\mathbf{s}) = 1. \tag{50}$$

Furthermore, from Proposition 5(c), we know that $\tilde{\mathbf{z}}_C \sim \tilde{d} \cdot \tilde{\mathbf{s}}(1)$ with $\tilde{d} \sim \|\tilde{\mathbf{z}}_C\|$. Thus,

$$\mathbb{P}\Big[\mathscr{E}_i[\tilde{\boldsymbol{z}}_C] \Big| \|\tilde{\boldsymbol{z}}_C\| \leq r_A\Big] = \mathbb{P}\Big[\mathscr{E}_i\Big[\tilde{d} \cdot \tilde{\mathbf{s}}(1)\Big] \Big| \tilde{d} \leq r_A\Big]. \tag{51}$$

Next observe that, conditioned on $\tilde{d} \leq r_A$, the event $\mathscr{E}_i\left[\tilde{d}\cdot\tilde{\mathbf{s}}\,(1)\right]$ implies that $f_i\left(\tilde{\mathbf{s}}\,(1)\right)=1$, leading to

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{d}\tilde{\mathbf{s}}\,(1)\right]\Big|\tilde{d}\leq r_A\right]\leq\mathbb{P}\left[f_i\left(\tilde{\mathbf{s}}\,(1)\right)=1\right]. \tag{52}$$

Finally, noting that conditioned on $d=r_A$, $\tilde{\mathbf{u}}_A\sim\tilde{\mathbf{s}}(1)$. Thus,

$$\begin{aligned}
\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{u}}_A\right]\right] &= \mathbb{P}_{\mathbf{s}\sim\tilde{\mathbf{s}}(1)}\left[\mathscr{E}_i\left[d\mathbf{s}\right]\Big|d=r_A\right]\\
&= \mathbb{P}_{\mathbf{s}\sim\tilde{\mathbf{s}}(1)}\left[g_i\left(\mathbf{s}\right)=1\right] && \text{(from (50))}\\
&\geq \mathbb{P}_{\mathbf{s}\sim\tilde{\mathbf{s}}(1)}\left[f_i\left(\mathbf{s}\right)=1\right] && \text{(from (49))}\\
&\geq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\boldsymbol{z}}_C\right]\Big|\left\|\tilde{\boldsymbol{z}}_C\right\|\leq r_A\right]. && \text{(from (51) and (52))}
\end{aligned}$$

(c) The proof is very similar to that of part (b) and is omitted. □

Proof of Proposition 7.
Applying Proposition 6(c) with $\tilde{\mathbf{u}}_A=\tilde{\mathbf{z}}_U$, $r_A=\overline{\sigma}\sqrt{n}$, and $\tilde{\mathbf{z}}_G=\tilde{\mathbf{z}}_C$, we obtain :

$$\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_U\right]\right]\;\leq\;\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\Big|\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]. \tag{53}$$

Furthermore,

$$\begin{aligned}
\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right] &= \mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|^2>n\overline{\sigma}^2\right]\\
&= \mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|^2>n\sigma^2-\left(n\sigma^2-n\overline{\sigma}^2\right)\right]\\
&= \mathbb{P}\left[\frac{1}{n}\left\|\tilde{\mathbf{z}}_G\right\|^2>\sigma^2-r\right],
\end{aligned}$$

where $r=\sigma^2-\overline{\sigma}^2$. Applying Proposition 5(a), we have

$$\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]\geq 1-\exp\left(-n\beta\right),\;\text{with }\beta=\frac{r-\log\left(1+r\right)}{2\sigma^2}. \tag{54}$$

Therefore, we have

$$\begin{aligned}
\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\right] &= \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\Big|\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]\cdot\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]\\
&\quad+\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\Big|\left\|\tilde{\mathbf{z}}_G\right\|\leq\overline{\sigma}\sqrt{n}\right]\cdot\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|\leq\overline{\sigma}\sqrt{n}\right]\\
&\geq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\Big|\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]\cdot\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]\\
&\geq \mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_U\right]\right]\cdot\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right], \quad \text{(from (53))}
\end{aligned}$$

which from (54) implies that $\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_U\right]\right]\leq\dfrac{\mathbb{P}\left[\mathscr{E}_i\left[\tilde{\mathbf{z}}_G\right]\right]}{\mathbb{P}\left[\left\|\tilde{\mathbf{z}}_G\right\|>\overline{\sigma}\sqrt{n}\right]}\;\leq\;\dfrac{\epsilon}{1-\exp\left(-n\beta\right)}$. □

Proof of Proposition 8.
(a) Constraint (23) implies that $\{\mathbf{x}_i\}_{i\in\mathcal{M}}$ satisfy that, for the noise vector $\mathbf{z}_t$'s in $\mathcal{Z}$ with $v_{it}=1$,

$$\begin{aligned}
\left\|\mathbf{x}_i-\mathbf{x}_k+\mathbf{z}_t\right\|\geq\left\|\mathbf{z}_t\right\|,\;\forall k\neq i,\;&\Longleftrightarrow\;\left\|\mathbf{x}_i-\mathbf{x}_k\right\|^2\geq 2\left\langle\mathbf{x}_k-\mathbf{x}_i\,,\,\mathbf{z}_t\right\rangle,\;\forall k\neq i,\\
&\Longleftrightarrow\;\left\|\mathbf{x}_i-\mathbf{x}_k\right\|^2\geq 2\left(1+\nu\right)\left\langle\mathbf{x}_k-\mathbf{x}_i\,,\,\mathbf{w}_t\right\rangle,\;\forall k\neq i.
\end{aligned} \tag{55}$$

Since $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T\}$ forms a Voronoi tessellation of $\mathcal{S}_n((1+\nu)\gamma_\epsilon)$, the set $\mathcal{W}$ also forms a Voronoi tessellation of $\mathcal{S}_n(\gamma_\epsilon)$. Therefore, letting $\mathbf{s}$ be any vector belonging to $\mathcal{S}_n(\gamma_\epsilon)$, from Proposition 4 applied to $\mathcal{A} = \mathcal{W}$, $N = T$, $\Lambda = \gamma_\epsilon/\sqrt{n}$, we obtain

$$\left\|\mathbf{s} - \mathbf{w}_{\tau(\tilde{\mathbf{s}})}\right\| \leq \theta\sqrt{n}, \tag{56}$$

where $\tau(\mathbf{s}) = \arg\min_{t\in\mathcal{T}} \|\mathbf{s} - \mathbf{w}_t\|$, with

$$\theta = \frac{\gamma_\epsilon}{T^{1/n}\cdot\sqrt{n}} = \frac{\zeta\nu}{1+\nu}. \tag{57}$$

From (25) and (57), we have

$$\|\mathbf{x}_i - \mathbf{x}_k\| \geq 2\sqrt{n}\zeta = 2\theta\frac{1+\nu}{\nu}\sqrt{n}. \tag{58}$$

We now proceed to show that if $v_{i\tau(\mathbf{s})} = 1$, then $\|\mathbf{x}_i - \mathbf{x}_k + \mathbf{s}\| \geq \|\mathbf{s}\|$. We have

$$\begin{aligned}
2\left\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \mathbf{s}\right\rangle &= 2\left\langle \mathbf{x}_k - \mathbf{x}_i, \left(\mathbf{s} - \mathbf{w}_{\tau(\mathbf{s})}\right) + \mathbf{w}_{\tau(\mathbf{s})}\right\rangle \\
&\leq 2\left\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \mathbf{w}_{\tau(\mathbf{s})}\right\rangle + 2\cdot\|\mathbf{x}_i - \mathbf{x}_k\|\cdot\left\|\mathbf{s} - \mathbf{w}_{\tau(\mathbf{s})}\right\| \quad \text{(Cauchy-Schwartz)} \\
&\leq (1+\nu)^{-1}\|\mathbf{x}_i - \mathbf{x}_k\|^2 + 2\cdot\|\mathbf{x}_i - \mathbf{x}_k\|\cdot\theta\sqrt{n} \quad \text{(from (55) and (56))} \\
&= \|\mathbf{x}_i - \mathbf{x}_k\|\cdot\left\{(1+\nu)^{-1}\|\mathbf{x}_i - \mathbf{x}_k\| + 2\theta\sqrt{n}\right\} \\
&\leq \|\mathbf{x}_i - \mathbf{x}_k\|\cdot\left\{(1+\nu)^{-1}\|\mathbf{x}_i - \mathbf{x}_k\| + \frac{\nu}{1+\nu}\|\mathbf{x}_i - \mathbf{x}_k\|\right\} \quad \text{(from (58))} \\
&= \|\mathbf{x}_i - \mathbf{x}_k\|^2.
\end{aligned}$$

Therefore, if $v_{i\tau(\mathbf{s})} = 1$, then

$$2\left\langle \mathbf{x}_k - \mathbf{x}_i\,,\, \mathbf{s}\right\rangle \leq \|\mathbf{x}_i - \mathbf{x}_k\|^2 \iff \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{s}\| \geq \|\mathbf{s}\|.$$

This implies that $\mathbb{P}\left[\forall k \neq i\,:\, \|\mathbf{x}_i - \mathbf{x}_k + \tilde{\mathbf{s}}\| \geq \gamma_\epsilon\right] \geq \mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\right]$.
(b) For each $i \in \mathcal{B}$, we have

$$\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\right] = \sum_{t=1}^{T}\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\middle|\tau(\tilde{\mathbf{s}}) = t\right]\cdot\mathbb{P}\left[\tau(\tilde{\mathbf{s}}) = t\right]. \tag{59}$$

We have

$$\begin{aligned}
\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})} = 1\middle|\tau(\tilde{\mathbf{s}}) = t\right] &= \begin{cases} 1, & \text{if } v_{it} = 1, \\ 0, & \text{otherwise,}\end{cases} \\
&= v_{it}.
\end{aligned} \tag{60}$$

Moreover, since the set of vectors $\mathcal{W}$ forms a Voronoi tessellation of $\mathcal{S}_n(\gamma_\epsilon)$, the Voronoi regions of the points $\mathbf{w}_t \in \mathcal{W}$ are identical with the same area. Consequently, choosing $\tilde{\mathbf{s}}$ uniformly induces a uniform distribution for $\tau(\tilde{\mathbf{s}})$ on the elements of the set $\mathcal{T}$, that is,

$$\mathbb{P}\left[\tau(\tilde{\mathbf{s}}) = t\right] = \frac{1}{T},\ \forall t \in \mathcal{T}. \tag{61}$$

Substituting (60) and (61) in (59), we have

$$\begin{aligned}\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})}=1\right] &= \sum_{t=1}^{T}\mathbb{P}\left[v_{i\tau(\tilde{\mathbf{s}})}=1\middle|\tau\left(\tilde{\mathbf{s}}\right)=t\right]\cdot\mathbb{P}\left[\tau\left(\tilde{\mathbf{s}}\right)=t\right]\\ &= \frac{1}{T}\sum_{t=1}^{T}v_{it}\geq\frac{1}{T}\cdot\left(1-\epsilon\right)T \qquad \text{(from (24))}\\ &= 1-\epsilon.\end{aligned}$$

□

Appendix B - Computing vectors $\mathcal{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_T\}$ using Lloyd's Algorithm

In this section, we present Lloyd's algorithm (Lloyd [1982]) to compute the vectors $\mathcal{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_T\}$.

Algorithm 4. ***Lloyd's Algorithm (Lloyd [1982])***
Input: *Sphere $\mathcal{B}(r)$, parameters δ_0, T.*
Output: *A set of vectors $\mathcal{Z}=\{\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_T\}$ that form a Voronoi tessellation of $\mathcal{B}(r)$ with T points.*
Algorithm:

1. *Generate a uniform distribution of vectors $\mathcal{Z}^0=\{\mathbf{z}_1^0,\mathbf{z}_2^0,\ldots,\mathbf{z}_T^0\}$ over $\mathcal{B}(r)$. Set $k=0$.*
2. *Repeat*
   - *Compute the Voronoi diagram of the points in the set $\mathcal{Z}^k$.*
   - *Let $\mathbf{z}_i^{k+1}$ be the centroid of the Voronoi cell that point $\mathbf{z}_i^k$ belongs to.*
   - *If $\max_{i=1,\ldots,T}\|\mathbf{z}_i^{k+1}-\mathbf{z}_i^k\|\leq\delta_0$, stop and output $\mathcal{Z}^{k+1}$; otherwise set $k:=k+1$ and go to Step 2.*

Sabin and Gray [1986] have shown that Algorithm 4 converges to a Voronoi tessellation of $\mathcal{B}(r)$ with T points for appropriately chosen values of δ_0 (see also Du et al. [1999, 2006]). Liu et al. [2009] have also shown that this algorithm converges at a linear rate, and have proposed other algorithm that converge faster. Note also that each iteration takes $\mathcal{O}(nT)$ steps. In the context of information theory Gray and Neuhoff [1998] discuss applications of Algorithm 4 to generate rate-distortion codebooks.