Title
Convex geometric tools in information theory

Author
Jog, Varun S.

Publication Date
2015

Convex Geometric Tools in Information Theory

by

Varun Suhas Jog

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering — Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Venkat Anantharam, Chair
Professor Martin Wainwright
Professor Aditya Guntuboyina

Summer 2015

Convex Geometric Tools in Information Theory

Abstract

Convex Geometric Tools in Information Theory

by

Varun Suhas Jog

Doctor of Philosophy in Engineering — Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Venkat Anantharam, Chair

The areas of information theory and geometry mirror each other in remarkable ways, with several concepts in geometry having analogues in information theory. These observations provide a simple way to posit theorems in one area by translating the corresponding theorems in the other. However, the analogy does not extended fully, and the proof techniques often do not carry over without substantial modification. One reason for this is that information theoretic quantities are often defined asymptotically, as the dimension tends to infinity. This is in contrast to the setting in geometry, where the dimension is usually fixed. In this dissertation, we try to bridge the gap between these two areas by studying the asymptotic geometric properties of sequences of sets. Our main contribution is developing a theory to study the growth rates of intrinsic volumes for sequences of convex sets satisfying some natural growth constraints. As an illustration of the usefulness of our techniques, we consider two specific problems. The first problem is that of analyzing the Shannon capacities of power-constrained communication channels. In particular, we study a power-constrained channel arising out of the energy harvesting communication model, called the (σ, ρ)-power constrained additive white Gaussian noise (AWGN) channel. Our second problem deals with forging new connections between geometry and information theory by studying the intrinsic volumes of sequences of typical sets. For log-concave distributions, we show the existence of a new quantity called the intrinsic entropy, which can be interpreted as a generalization of differential entropy.

To Aai, Baba, and Amod

Contents

List of Figures

Acknowledgments

My experience over the past five years at Berkeley can be best described as a roller-coaster ride, played in super slow motion. The ride was replete with thrilling highs, seemingly never-ending lows, and some surprising twists and turns. Just like a roller-coaster ride, the dominant emotion at the end is relief at having survived, and eager anticipation for the next ride!

I was lucky to have interacted with many amazing people at Berkeley, whose friendship and guidance enriched my life here. They helped me mature as an individual, and as a researcher. I would like to thank all of them.

The largest debt of gratitude is owed to my advisor, Venkat Anantharam. Venkat gave me considerable freedom to choose problems to work on, and played a very active role in guiding me to solve them. He invested countless hours carefully reading my manuscripts (including this one) and fixing its technical as well as grammatical errors. I especially appreciate how accessible he was throughout my PhD, whether it was scheduling in-person meetings, responding to e-mails, or Skyping during travel. Caring, cultured, and highly knowledgeable, Venkat is a role model for the kind of person I would like to be.

I am grateful to Martin Wainwright and Aditya Guntuboyina for being a part of my dissertation committee, and Anant Sahai for being a part of my qualifying exam committee. I would like to thank Tom Courtade and Abhay Parekh for giving me the opportunity to be a GSI for my favorite courses, EE226A and EE126A. Aditya and Tom were very supportive in my postdoc search, and for this I am grateful.

I want to thank the EECS administrative staff, especially Shirley Salanio and Kim Kail, for being extremely reliable and efficient in resolving all my administrative issues.

Wireless Foundations has been a terrific place to hang out during these past years. Coming to lab was always something to look forward to, thanks to my labmates Po-Ling Loh, Vijay Kamble, Kangwook Lee, Giulia Fanti, Stephan Adams, Nihar Shah, Rashmi K.V., Ramtin Pedarsani, Ka Kit Lam, Vasuki Narasimha Swamy, Govinda Kamath, Steven Clarkson, Venky Ekambaram, Sameer Pawar, Naveen Goela, Gireeja Ranade, Se Yong Park, Kate Harrison, Sudeep Kamath, Fanny Yang, Reza Abbasi Asl, Ashwin Pananjady, Vidya Muthukumar, Orhan Ocal, Dong Yin, and Payam Delgosha. A special thanks to my roommates over the years, Shashank Nawathe and Saurabh Gupta, who made going home from lab something to look forward to as well.

Life in Berkeley would not have been so memorable had it not been for my biking partner, research collaborator, and fiancée, Po-Ling Loh.

Lastly, I would like to thank my family — my mother, Kalpana Jog, my father, Suhas Jog, and my brother and best friend, Amod Jog. Their love and support means the world to me, and this dissertation is dedicated to them.

Chapter 1

Introduction

Concepts in geometry often have parallels in information theory; for example, volume and entropy, surface area and Fisher information, sphere-packing and channel coding, and Euclidean balls and Gaussian distributions, to name a few. This connection is perhaps best exemplified by a striking similarity between two fundamental inequalities in the fields:

- Entropy power inequality (EPI) [33]: For independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ over $\mathbb{R}^n$,
$$e^{\frac{2}{n}h(\mathbf{X})} + e^{\frac{2}{n}h(\mathbf{Y})} \leq e^{\frac{2}{n}h(\mathbf{X}+\mathbf{Y})}.$$
- Brunn-Minkowski inequality (BMI) [14]: For convex sets $A, B \subseteq \mathbb{R}^n$,
$$\mathrm{Vol}(A)^{1/n} + \mathrm{Vol}(B)^{1/n} \leq \mathrm{Vol}(A \oplus B)^{1/n},$$
where $A \oplus B$ denotes the Minkowski sum of A and B.

Further similarities can be found in the area of isoperimetric inequalities. In geometry, the isoperimetric inequality states that amongst all bodies having a fixed volume the Euclidean ball has the least surface area. The analogue of this is the entropic isoperimetric inequality which states that among all distributions with a fixed entropy, the Gaussian distribution has the least Fisher information. Yet another example is that of Costa's EPI [6] in information theory which implies the *concavity of entropy power* described as follows: For an $\mathbb{R}^n$ valued random variable X and white Gaussian $Z \sim \mathcal{N}(0, I)$, the function

$$h(t) = \exp\left(\frac{2}{n}h(X + \sqrt{t}Z)\right)$$

is concave in t. Costa and Cover [5] observed an analogous concavity of the normalized volume function

$$v(t) = |A \oplus tB|^{1/n}$$

where $A \subseteq \mathbb{R}^n$ is a compact convex set and B is the Euclidean unit ball in $\mathbb{R}^n$.

Such similarities between geometry and information theory provide a simple way to posit theorems in one area by translating the corresponding theorems in the other. Our starting point in this dissertation is a sequence of sets, $\{K_n \subset \mathbb{R}^n : n \geq 1\}$. Although geometry normally deals with sets in a fixed dimension, such sequences show up naturally in information theory. For example, K_n may be the typical set of a probability distribution in dimension n, or K_n may be the set of all allowable codewords of length n for an input-constrained communication channel. When it exists, the growth rate of volume given by $v := \lim_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(K_n)$, is a very useful quantity in information theory. When K_n is a sequence of typical sets, v equals the differential entropy of the distribution. When K_n is the sequence of allowed codewords for a power-constrained channel coding, v can lead to asymptotically tight bounds on channel capacity. This gives rise to the following question: Are there are any other "useful" properties, besides volume?

In this dissertation, we study a class of geometric properties called *intrinsic volumes*. Intrinsic volumes are functions defined on the class of compact convex sets, and can be uniquely extended to *polyconvex sets*; i.e., sets which are finite unions of compact convex sets. We denote the set of compact convex sets in $\mathbb{R}^n$ by $\mathcal{C}_n$. A set $K \in \mathcal{C}_n$ has $n+1$ intrinsic volumes which are denoted by $\{V_0(K), \ldots, V_n(K)\}$. Some of these intrinsic volumes are known in the literature under alternate names; e.g. $V_0(K)$ is the Euler characteristic, $V_1(K)$ is the mean width, $V_{n-1}(K)$ is proportional to the surface area, and $V_n(K)$ is the volume. Intrinsic volumes have a number of interpretations in geometry. We state some of these interpretations as found in Klain & Rota [20]. Intrinsic volumes are *valuations* on $\mathcal{C}_n$; i.e. for all $A, B \in \mathcal{C}_n$ such that $A \cup B \in \mathcal{C}_n$, and for all $0 \leq i \leq n$,

$$V_i(A \cup B) = V_i(A) + V_i(B) - V_i(A \cap B).$$

Furthermore, these valuations are convex-continuous [30] and invariant under rigid motions. In fact, Hadwiger's theorem states that any convex-continuous, rigid-motion invariant valuation on $\mathcal{C}_n$ has to be a linear combination of the intrinsic volume valuations. Here convex-continuity is with respect to the topology on $\mathcal{C}_n$ induced by the Hausdorff metric δ, which gives the distance between $A, B \in \mathcal{C}_n$ by the relation

$$\delta(A, B) = \max(\sup_{a\in A} \inf_{b\in B} |a-b|, \sup_{b\in B} \inf_{a\in A} |a-b|).$$

Kubota's theorem or Crofton's formula implies that the i^{th} intrinsic volume $V_i(K)$ is proportional to the volume of a random i-dimensional projection or slice of K. Intrinsic volumes are thus defined by the geometric structure of a set and describe its global characteristics.

Given a sequence $\mathcal{K} = \{K_n\}_{n\geq 1}$ such that $K_n \in \mathcal{C}_n$, our primary goal is to identify the growth rate of intrinsic volumes for this sequence. To be precise, for $\theta \in [0, 1]$ we want to study the existence of the limit

$$\mathcal{G}_{\mathcal{K}}(\theta) = \lim_{n\to\infty} \frac{1}{n} \log V_{\lfloor n\theta \rfloor}(K_n). \tag{1.1}$$

The value $\mathcal{G}_{\mathcal{K}}(\theta)$ gives the growth rate of the $n\theta^{\text{th}}$ intrinsic volume of the sequence $\mathcal{K}$. We call $\mathcal{G}_{\mathcal{K}}$ the *growth function* of $\mathcal{K}$, abbreviated as the $\mathcal{G}$-function of $\mathcal{K}$. As an illustration, we compute the $\mathcal{G}$-function of the sequence for cubes and the sequence of balls:

Example 1. Consider the sequence of cubes $K_n = [0, A]^n$ for some $A > 0$. The intrinsic volumes of cubes are known in a closed form [20],

$$V_i(K_n) = \binom{n}{i} A^i. \tag{1.2}$$

Taking the appropriate limits, we see that

$$\mathcal{G}_{\mathcal{K}}(\theta) = H(\theta) + \log A, \tag{1.3}$$

where $H(\theta) = -\theta \log \theta - (1-\theta)\log(1-\theta)$[1] is the binary entropy function.

Example 2. Consider the sequence of Euclidean balls whose radii grow linearly with the dimension; i.e. $K_n = B_n(\sqrt{n\nu})$ for some $\nu > 0$. The i^{th} intrinsic volume of K_n is given by [20]

$$V_i(K_n) = \binom{n}{i} \frac{\omega_i}{\omega_{n-i}} (n\nu)^{i/2}, \tag{1.4}$$

where ω_i is the volume of the i-dimensional unit ball. Taking the appropriate limits, we obtain the $\mathcal{G}$-function to be

$$\mathcal{G}_{\mathcal{K}}(\theta) = H(\theta) + \frac{\theta}{2} \log 2\pi e\nu + \frac{1-\theta}{2} \log(1-\theta). \tag{1.5}$$

In general, evaluating or even showing the existence of $\mathcal{G}_{\mathcal{K}}$ is challenging because intrinsic volumes are notoriously hard to compute, and rarely available in closed form. We therefore have to identify specific structural properties of $\mathcal{K}$ and use these to establish the existence of a growth rate. In this dissertation, we identify two such properties which enable us to show the existence of a growth rate for a large class of sequences.

- **Sub-convolutive sets:** For all $m, n \geq 1$, the sequence $\mathcal{K}$ satisfies

$$K_{m+n} \subseteq K_m \times K_n. \tag{1.6}$$

- **Super-convolutive sets:** For all $m, n \geq 1$, the sequence $\mathcal{K}$ satisfies

$$K_m \times K_n \subseteq K_{m+n}. \tag{1.7}$$

[1] All logarithms are to base e.

Sub/super convolutive sequences of sets show up naturally in the context of power-constrained channels in information theory when describing whether or not the concatenation of two valid codewords x^n and y^m is a valid codeword. The amplitude constraint or the average power constraint are examples of super-convolutive constraints. In Chapter 3 we will encounter energy harvesting communication systems, where the transmitter is subject to a sub-convolutive power constraint called the (σ, ρ)-power constraint.

To simplify notation, we denote the $n+1$ intrinsic volumes of K_n by $\mu_n(0), \ldots, \mu_n(n)$. The values of $\mu_n(i)$ for $i > n$ are taken to be 0. Thus, μ_n can be thought of as a function from $\mathbb{Z}_+ \to \mathbb{R}_+$. We use the following properties of intrinsic volumes [20]: For all $A, B \in \mathcal{C}^n$,

- If $A \subseteq B$ then $V_j(A) \leq V_j(B)$ for all $0 \leq j \leq n$; i.e., intrinsic volumes are monotonic with respect to inclusion.
- For $A, B \in \mathcal{C}_n$, the intrinsic volumes of $A \times B$ are obtained by convolving the intrinsic volumes of A and B; i.e., for all $j \geq 0$,

  $$V_j(A \times B) = \sum_{i=0}^{\infty} V_i(A) V_{j-i}(B). \tag{1.8}$$

  Here, we use the fact that $V_i(A) = V_i(B) = 0$ for $i > n$.

Using these, it is easy to see that the intrinsic volumes of sub/super-convolutive sets must satisfy:

- **Sub-convolutive sequence:** If $\mathcal{K}$ is sub-convolutive, then for all $m, n \geq 1$, the intrinsic volume sequences satisfy

  $$\mu_{m+n} \leq \mu_m \star \mu_n, \tag{1.9}$$

  where $\mu_m \star \mu_n$ is the convolution of the sequences of intrinsic volumes of K_m and K_n respectively.

- **Super-convolutive sequences:** If $\mathcal{K}$ is super-convolutive, then for all $m, n \geq 1$, the intrinsic volume sequences satisfy

  $$\mu_m \star \mu_n \leq \mu_{m+n}, \tag{1.10}$$

  where $\mu_m \star \mu_n$ is the convolution of the sequences of intrinsic volumes of K_m and K_n respectively.

Starting from this simple observation, our research builds a framework to analyze intrinsic volumes of sub/super convolutive sequences of sets. An outline of this thesis is as follows:

- In Chapter 2, we study the convergence properties of sub-convolutive and super-convolutive sequences. The theory of large deviations plays a key role in the analysis of such sequences.

- In Chapter 3, we consider the problem of finding the capacity of a power-constrained additive white Gaussian noise (AWGN) channel. The power-constraints, called (σ, ρ)-constraints, are motivated by energy harvesting communication systems. We show that the sequence of sets describing all the allowable length n sequences forms a *sub-convolutive sequence*. Among other results, a notable contribution of this chapter is establishing an high-dimensional version of *Steiner's formula* from convex geometry, which describes the volume of the Minkowski sum of a convex set and a ball. The main results in this chapter have previously appeared in the publications [17] and [16].

- In Chapter 4, we observe that typical sets of log-concave distributions form a super-convolutive sequence. Using this, we establish the existence of a quantity called "intrinsic entropy" which is a generalization of the notion of differential entropy. This chapter contains results which were previously published in [18].

- We conclude with Chapter 5 where we discuss some open problems and future directions.

Chapter 2

Sub and super-convolutive sequences

As described in Chapter 1, sub and super-convolutive sequences emerge naturally in the study of intrinsic volumes of sequences of sets. We define such sequences more precisely as follows. Consider a sequence of functions $\{\mu_n(\cdot)\}_{n\geq 1}$, such that for every n, $\mu_n : \mathbb{Z}_+ \to \mathbb{R}_+$ with $\mu_n(j) = 0$ for all $j \geq n+1$. We call such a sequence of functions a *sub-convolutive* sequence if for all $m, n \geq 1$ the convolution $\mu_m \star \mu_n$ pointwise dominates μ_{m+n}; i.e.,

$$\mu_m \star \mu_n(i) \geq \mu_{m+n}(i) \text{ for all } i \geq 0, \text{ and for all } m, n \geq 1. \tag{2.1}$$

Similarly, we call such a sequence of functions a *super-convolutive* sequence if for all $m, n \geq 1$ the convolution $\mu_m \star \mu_n$ is pointwise dominated by μ_{m+n}; i.e.,

$$\mu_m \star \mu_n(i) \leq \mu_{m+n}(i) \text{ for all } i \geq 0, \text{ and for all } m, n \geq 1. \tag{2.2}$$

Such sequences can be effectively studied using results from the theory of large deviation, and in particular the Gärtner-Ellis theorem [10] stated here:

Theorem 2.0.1 (Gärtner-Ellis theorem)**.** *Consider a sequence of random vectors $Z_n \in \mathbb{R}^d$, where Z_n possess the law ν_n and the logarithmic moment generating function*

$$\Lambda_n(\lambda) := \log E\left[\exp\langle \lambda, Z_n\rangle\right].$$

We assume the following:

($\star$)*: For each $\lambda \in \mathbb{R}^d$, the logarithmic moment generating function, defined as the limit*

$$\Lambda(\lambda) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$$

exists as an extended a real number. Further the origin belongs to the interior $\mathcal{D}_\Lambda := \{\lambda \in \mathbb{R}^d \mid \Lambda(\lambda) < \infty\}$.

Let Λ^* be the convex conjugate of λ with $\mathcal{D}_{\Lambda^*} = \{x \in \mathbb{R}^d \mid \Lambda^*(x) < \infty\}$. When assumption $(\star)$ holds, the following are satisfied:

1. *For any closed set I,*
$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n(I) \leq -\inf_{x\in I} \Lambda^*(x).$$

2. *For any open set F,*
$$\liminf_{n\to\infty} \frac{1}{n} \log \nu_n(F) \geq -\inf_{x\in F\cap\mathcal{F}} \Lambda^*(x),$$
where $\mathcal{F}$ is the set of exposed points of Λ^* whose exposing hyperplane belongs to the interior of $\mathcal{D}_\Lambda$.

3. *If Λ is an essentially smooth, lower semicontinuous function, then the large deviations principle holds with a good rate function Λ^*.*

Remark 2.0.2. For definitions of *exposed points*, *essentially smooth functions*, *good rate function*, and the *large deviations principle* we refer to Section 2.3 of [10]. For our purpose, it is enough to know that if Λ is differentiable on $\mathcal{D}_\Lambda = \mathbb{R}^d$, then it is essentially smooth and Λ^* satisfies the large deviation principle.

2.1 Convergence properties of sub-convolutive sequences

For our results on sub-convolutive sequences, we make the following assumptions:

$$(\mathbf{A}) : \alpha := \lim_{n\to\infty} \frac{1}{n} \log \mu_n(n) \quad \text{is finite.}$$
$$(\mathbf{B}) : \beta := \lim_{n\to\infty} \frac{1}{n} \log \mu_n(0) \quad \text{is finite.}$$
$$(\mathbf{C}) : \text{For all } n,\ \mu_n(n) > 0, \mu_n(0) > 0.$$

Note that $\mu_m \star \mu_n(m+n) = \mu_n(n)\mu_m(m)$ and $\mu_m \star \mu_n(0) = \mu_n(0)\mu_m(0)$. Thus, the existence of the limits in assumptions $(\mathbf{A})$ and $(\mathbf{B})$ is guaranteed by Fekete's Lemma [35], and we have

$$\alpha = \inf_n \frac{1}{n} \log \mu_n(n) \tag{2.3}$$
$$\beta = \inf_n \frac{1}{n} \log \mu_n(0). \tag{2.4}$$

For $n \geq 1$, define $G_n : \mathbb{R} \to \mathbb{R}$ as

$$G_n(t) = \log \sum_{j=0}^{n} \mu_n(j) e^{jt}. \tag{2.5}$$

Condition (2.1) implies that the functions G_n satisfy the inequality,

$$G_m(t) + G_n(t) \geq G_{m+n}(t) \text{ for every } m, n \geq 1 \text{ and for every } t. \tag{2.6}$$

Thus for each t, the sequence $\{G_n(t)\}$ is sub additive, and by Fekete's lemma the limit $\lim_n \frac{G_n(t)}{n}$ exists. To simply notation a bit, define $g_n := \frac{G_n}{n}$ and let Λ be defined as the pointwise limit of g_n's; i.e.,

$$\Lambda(t) = \lim_n g_n(t). \tag{2.7}$$

Lemma 2.1.1 (Proof in Appendix B.1.1)**.** *The function Λ satisfies the following properties:*

1. *For all t,*
$$\max(\beta, t + \alpha) \leq \Lambda(t) \leq g_1(t) \tag{2.8}$$
2. *Λ is convex and monotonically increasing.*
3. *Let Λ^* be the convex conjugate of Λ. The domain of Λ^* is $[0, 1]$.*

Lemma 2.1.1 along with Theorem 2.0.1 lead to the following large deviations "upper bound" result:

Theorem 2.1.2. *Consider a sequence of functions $\{\mu_n(\cdot)\}_{n\geq 1}$, such that for every n, $\mu_n : \mathbb{Z}_+ \to \mathbb{R}_+$ with $\mu_n(j) = 0$ for all $j \geq n + 1$. Suppose $\{\mu_n\}_{n\geq 1}$ is a sequence of sub-convolutive functions as defined in equation (2.1), satisfying assumptions (**A**), (**B**) and (**C**). Define a sequence of measures supported on $[0, 1]$ by*

$$\mu_{n/n}\left(\frac{j}{n}\right) := \mu_n(j) \quad \textit{for} \quad j \geq 0.$$

Let $I \subseteq \mathbb{R}$ be a closed set. The family of measures $\{\mu_{n/n}\}$ satisfies the large deviation upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(I) \leq - \inf_{x \in I} \Lambda^*(x). \tag{2.9}$$

Proof. Let $\sum_j \mu_n(j) = s_n$. We first normalize $\mu_{n/n}$ to define the probability measure

$$p_n := \frac{\mu_{n/n}}{s_n}.$$

The log moment generating function of p_n, which we call P_n, is given by

$$\begin{aligned}
P_n(t) &= \log \sum_{j=0}^{n} p_n(j/n) e^{jt/n} \\
&= \log \frac{1}{s_n} \sum_{j=0}^{n} \mu_n(j) e^{jt/n} \\
&= G_n(t/n) - \log s_n.
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} P_n(nt) &= \lim_{n\to\infty} \left(\frac{G_n(t)}{n} - \frac{\log s_n}{n} \right) \\
&= \Lambda(t) - \Lambda(0).
\end{aligned}$$

Note also that by Lemma 2.1.1, the function Λ is finite on all of $\mathbb{R}$, and thus 0 lies in the interior $\mathcal{D}(\Lambda)$. Thus, the sequence of probability measures $\{p_n\}$ satisfies the condition $(\star)$ required in the Gärtner-Ellis theorem. A direct application of this theorem gives the bound

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log p_n(I) &\leq -\inf_{x\in I} (\Lambda(x) - \Lambda(0))^* \\
&= -\inf_{x\in I} \Lambda^*(x) - \Lambda(0),
\end{aligned}$$

which immediately gives

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(I) \leq -\inf_{x\in I} \Lambda^*(x).$$

$\square$

Remark 2.1.3. If $\Lambda(t)$ is differentiable, we can apply the Gärtner-Ellis theorem to get a lower bound of the form

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) \geq -\inf_{x\in F} \Lambda^*(x),$$

for every open set F. However, it is easy to construct sub-convolutive sequences such that $\Lambda(t)$ is not differentiable. One example is the sequence $\{\mu_n\}$ such that for each n,

$$\mu_n(j) = \begin{cases} 1 & \text{if } j = 0 \text{ or } n, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.1.4. *The functions $\{g_n^*\}$ converge uniformly to Λ^* on $[0,1]$.*

Proof. We'll show that $\{g_n^*\}$ converge pointwise to Λ^* on $[0,1]$. Since g_n^* and Λ^* are all continuous convex functions on a compact set, Lemma A.1.1 implies that this pointwise convergence implies uniform convergence.

Recall that $\alpha = \inf_n \frac{1}{n}\log \mu_n(n)$, $\beta = \inf_n \frac{1}{n}\log \mu_n(0)$, and $\inf_n g_n(0) = \Lambda(0)$. Fix an $x \in (0,1)$, and define

$$\arg\max_t xt - g_n(t) := t_n.$$

Clearly, $g_n^*(x) = xt_n - g_n(t_n)$. Note that

$$g_n^*(x) \geq xt - g_n(t)\Big|_{t=0} \tag{2.10}$$
$$= -g_n(0) \tag{2.11}$$
$$\overset{(a)}{\geq} -g_1(0) \tag{2.12}$$

where (a) follows by inequality (B.1).

If $t > \frac{g_1(0)-\alpha}{1-x}$, then we have

$$\begin{aligned} xt - g_n(t) &< xt - (t+\alpha) \\ &= -(1-x)t - \alpha \\ &< -(g_1(0)-\alpha) - \alpha \\ &= -g_1(0). \end{aligned}$$

This gives us that $t_n \leq \frac{g_1(0)-\alpha}{1-x}$. Similarly, if $t < \frac{\beta - g_1(0)}{x}$, then

$$xt - g_n(t) < xt - \beta \tag{2.13}$$
$$< (\beta - g_1(0)) - \beta \tag{2.14}$$
$$= -g_1(0). \tag{2.15}$$

This gives us that $t_n \geq \frac{\beta - g_1(0)}{x}$. We can thus conclude that for all n,

$$t_n \in \left[\frac{\beta - g_1(0)}{x}, \frac{g_1(0)-\alpha}{1-x}\right] := I_x. \tag{2.16}$$

Note that all we used to prove relation (2.16) is that g_n is trapped between g_1 and $\max(\beta, t+\alpha)$. Since Λ also satisfies this, we have

$$\arg\max_t xt - \Lambda(t) \in I_x. \tag{2.17}$$

We now restrict our attention to the compact interval I_x. Let $\hat{g}_n$ be g_n restricted to I_x. The convex functions $\hat{g}_n$ converge pointwise to a continuous limit $\hat{\Lambda}$, where $\hat{\Lambda}$ is Λ restricted to I_x. This convergence must therefore be uniform, which implies convergence of $\hat{g}_n^*(x)$ to $\hat{\Lambda}^*(x)$. Furthermore, relation (2.16) implies $\hat{g}_n^*(x)$ equals $g_n^*(x)$, and relation (2.17) gives $\hat{\Lambda}^*(x)$ equals $\Lambda^*(x)$. Thus, $g_n^*(\cdot)$ converges pointwise to $\Lambda^*(\cdot)$ on $(0,1)$.

We'll now consider convergence at the boundary points. Let $\epsilon > 0$ be given. Choose a subsequence $\{g_{n_k}\}$ where $n_k = 2^k$. Using the condition in (2.6), it is clear that $\{g_{n_k}\}$ decrease monotonically and converge pointwise to Λ. Choose K_0 large enough such that for all $k > K_0$,

$$\frac{1}{n_k} \log \mu_{n_k}(0) - \beta < \epsilon/2. \tag{2.18}$$

Note that the left hand side is non-negative, and we need not use absolute values. Choose a T_0 such that for all $t < T_0$,

$$g_{n_{K_0}}(t) - \frac{1}{n_{K_0}} \log \mu_{n_{K_0}}(0) < \epsilon/2. \tag{2.19}$$

Now for all $k > K_0$ and all $t < T_0$, the following holds:

$$g_{n_k}(t) - \beta \leq g_{n_{K_0}}(t) - \beta < \frac{1}{n_{K_0}} \log \mu_{n_{K_0}}(0) + \epsilon/2 - \beta < \epsilon. \tag{2.20}$$

Taking the limit in k, we get that for all $t < T_0$,

$$\Lambda(t) - \beta \leq \epsilon, \tag{2.21}$$

this along with the lower bound $\Lambda(t) \geq \beta$ gives that for all $t < T_0$,

$$0 \leq \Lambda(t) - \beta \leq \epsilon.$$

We also have

$$\Lambda^*(0) = \sup_t -\Lambda(t) \tag{2.22}$$

$$= -\lim_{t \to -\infty} \Lambda(t), \tag{2.23}$$

which must equal $-\beta$. Since the limit of $g_n^*(0)$ is also $-\beta$, we have shown convergence of g_n^* to Λ at $t = 0$.

To show convergence at $t = 1$, we follow a similar strategy. Let $\{g_{n_k}\}$ be as before, and let $\epsilon > 0$ be given. We choose a K_1 such that for all $k > K_1$,

$$\frac{1}{n_k} \log \mu_{n_k}(n_k) - \alpha < \epsilon/2. \tag{2.24}$$

Note that the left hand side is non-negative, and we need not use absolute values. We now choose a T_1 such that for all $t > T_1$,

$$g_{n_{K_1}}(t) - \left(t + \frac{1}{n_{K_1}} \log \mu_{n_{K_1}}(n_{K_1})\right) < \epsilon/2. \tag{2.25}$$

Now for all $k > K_1$ and all $t > T_1$,

$$g_{n_k}(t) - (t+\alpha) \leq g_{n_{K_1}}(t) - (t+\alpha) < \frac{1}{n_{K_1}} \log \mu_{n_{K_1}}(n_{K_1}) + \epsilon/2 - \alpha < \epsilon. \tag{2.26}$$

Taking the limit in k, we get that for all $t > T_1$,

$$\Lambda(t) - (t+\alpha) \leq \epsilon,$$

this along with the lower bound $\Lambda(t) \geq t + \alpha$ gives that for all $t > T_1$

$$0 \leq \Lambda(t) - (t+\alpha) \leq \epsilon.$$

From this, we conclude that $\Lambda^*(1) = \sup_t t - \Lambda(t) = \lim_{t\to+\infty} t - \Lambda(t)$, must equal $-\alpha$. Since the limit of $g_n^*(1)$ is also $-\alpha$, we have shown convergence of g_n^* to Λ at $t = 1$.

This shows that $\{g_n^*\}$ converges pointwise to Λ^* on the compact interval $[0, 1]$. As all the functions involved are continuous and convex, by Lemma A.1.1 this convergence must also be uniform. This concludes the proof. □

2.2 Convergence properties of super-convolutive sequences

Just as in the case of sub-convolutive sequences, we make certain assumption which the super-convolutive sequence $\{\mu_n\}$ should satisfy:

$$(\mathbf{A}) : \alpha := \lim_{n\to\infty} \frac{1}{n} \log \mu_n(n) \quad \text{is finite.}$$

$$(\mathbf{B}) : \beta := \lim_{n\to\infty} \frac{1}{n} \log \mu_n(0) \quad \text{is finite.}$$

$$(\mathbf{C}) : \text{For all } n,\ \mu_n(n) > 0, \mu_n(0) > 0.$$

Note that $\mu_m \star \mu_n(m+n) = \mu_n(n)\mu_m(m)$ and $\mu_m \star \mu_n(0) = \mu_n(0)\mu_m(0)$. Thus the existence of the limits in assumptions (**A**) and (**B**) is guaranteed by Fekete's Lemma, and we have

$$\alpha = \sup_n \frac{1}{n} \log \mu_n(n), \tag{2.27}$$

$$\beta = \sup_n \frac{1}{n} \log \mu_n(0). \tag{2.28}$$

For each $n \geq 1$, define $G_n : \mathbb{R} \to \mathbb{R}$ as

$$G_n(t) = \log \sum_{j=0}^{n} \mu_n(j) e^{jt}. \tag{2.29}$$

Condition (2.2) implies that the functions G_n satisfy the inequality,

$$G_m(t) + G_n(t) \leq G_{m+n}(t) \text{ for every } m, n \geq 1 \text{ and for every } t. \tag{2.30}$$

Thus for each t, the sequence $\{G_n(t)\}$ is super additive, and by Fekete's lemma the limit $\lim_n \frac{G_n(t)}{n}$ exists. Without further conditions on $\{\mu_n(\cdot)\}$, we cannot rule out this limit being $+\infty$ for some t. We therefore make an extra assumption, in addition to the assumptions $(\mathbf{A}), (\mathbf{B})$ and $(\mathbf{C})$.

$$(\mathbf{D}) : \gamma := \lim_{n \to \infty} \frac{G_n(0)}{n} \text{ is finite.}$$

To simply notation a bit, define $g_n := \frac{G_n}{n}$ and let Λ be defined as the pointwise limit of g_n's; i.e.,

$$\Lambda(t) = \lim_n g_n(t). \tag{2.31}$$

Lemma 2.2.1 (Proof in Appendix B.2.1)**.** *The function Λ satisfies the following properties:*

1. *For all t,*
$$g_1(t) \leq \Lambda(t) \leq \max(\gamma, t + \gamma) \tag{2.32}$$
2. *Λ is convex and monotonically increasing.*
3. *Let Λ^* be the convex conjugate of Λ. The domain of Λ^* is $[0, 1]$.*

Theorem 2.2.2. *Define a sequence of measures supported on $[0, 1]$ by*

$$\mu_{n/n}\left(\frac{j}{n}\right) := \mu_n(j) \quad \textit{for} \quad 0 \leq j \leq n.$$

Let $I \subseteq \mathbb{R}$ be a closed set. The family of measures $\{\mu_{n/n}\}$ satisfies the large deviation upper bound

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_{n/n}(I) \leq - \inf_{x \in I} \Lambda^*(x). \tag{2.33}$$

Proof. The proof is exactly same as the proof of Theorem 2.1.2. $\square$

Lemma 2.2.3 (Proof in Appendix B.2.2)**.** *Define $\Psi^* : [0, 1] \to \mathbb{R}$ to be the pointwise limit of g_n^*:*

$$\Psi^*(t) = \lim_{n \to \infty} g_n^*(t).$$

Then for $t \in (0, 1)$, we have $\Lambda^*(t) = \Psi^*(t)$, and for $t = 0$ and $t = 1$, we have $\Lambda^*(t) \leq \Psi^*(t)$.

Theorem 2.2.4. *Let $F \subseteq \mathbb{R}$ be an open set. The family of measures $\{\mu_{n/n}\}$ satisfies the large deviations lower bound*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) \geq -\inf_{x\in F} \Lambda^*(x). \tag{2.34}$$

Proof. We will construct a new sequence of functions $\{\hat{\mu}_n\}$ such that $\mu_n \geq \hat{\mu}_n$ for all n; i.e., μ_n pointwise dominates $\hat{\mu}_n$ for all n. The large deviations lower bound for the sequence $\{\hat{\mu}_n\}$ will then serve as a large deviations lower bound for the sequence $\{\mu_n\}$.

Fix an $a \geq 1$. We express every $n \geq 1$ as $n = qa + r$, where $r < a$, and define $\hat{\mu}_n = \mu_a^{\star q} \star \mu_r$. The super convolutive condition immediately implies

$$\mu_n \geq \hat{\mu}_n.$$

Define $\hat{G}_n(t)$ as follows,

$$\hat{G}_n(t) = \log \sum_{j=0}^{n} \hat{\mu}_n(j) e^{jt},$$

and consider the limit

$$\lim_{n\to\infty} \frac{1}{n} \hat{G}_n(t) = \lim_{n\to\infty} \frac{1}{n}(qG_a(t) + G_r(t)). \tag{2.35}$$

Note that the limit

$$\lim_{n\to\infty} \frac{1}{n}|G_r(t)| \leq \lim_{n\to\infty} \frac{1}{n} \max_{1\leq j\leq a-1} |G_j(t)| = 0.$$

Note also that $q = \lfloor n/a \rfloor$. Thus the limit in equation (2.35) evaluates to

$$\lim_{n\to\infty} \frac{1}{n} \lfloor n/a \rfloor G_a(t) = \frac{G_a(t)}{a} = g_a(t).$$

Applying Gärtner-Ellis theorem for $\{\hat{\mu}_n\}$, and noting that $g_a(t)$ is differentiable, we get the lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{\mu}_{n/n}(F) \geq -\inf_{x\in F} g_a^*(x), \tag{2.36}$$

which implies

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) \geq -\inf_{x\in F} g_a^*(x). \tag{2.37}$$

Taking the limit in a, and using Lemma A.2.1, we arrive at

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) &\geq -\inf_{x\in F} \Psi^*(x) \\ &= -\inf_{x\in F} \Lambda^*(x), \end{aligned}$$

which concludes the proof. $\square$

Remark 2.2.5. The assumptions $(\mathbf{A}), (\mathbf{B}), (\mathbf{C}), (\mathbf{D})$ do not ensure that Λ^* and Ψ^* agree at the boundary points 0 and 1. We give one example of a super-convolutive sequence where this disagreement occurs. Let $\epsilon, \alpha > 0$ such that $\alpha \geq 1, \epsilon < \frac{1}{2}$. Let's define a sequence of functions μ_n for $n \geq 1$ as follows:

$$\mu_n(i) = \begin{cases} \binom{n-1}{i}\alpha^i, & \text{for } 0 \leq i \leq n-1, \\ \epsilon & \text{for } i = n. \end{cases} \tag{2.38}$$

As shown in Appendix B.2.3, the above sequence is an example of a super-convolutive sequence satisfying the assumptions $(\mathbf{A}), (\mathbf{B}), (\mathbf{C}), (\mathbf{D})$ and yet having $\Lambda^*(1) \neq \Psi^*(1)$.

Chapter 3

The (σ, ρ)-power constrained AWGN channel

The additive white Gaussian noise (AWGN) channel is one of the most basic channel models studied in information theory. This channel is represented by a sequence of channel inputs denoted by X_i, and an input-independent additive noise Z_i. The noise variables Z_i are assumed to be independent and identically distributed as $\mathcal{N}(0, \nu)$. The channel output Y_i is given by

$$Y_i = X_i + Z_i \quad \text{for} \quad i \geq 1. \tag{3.1}$$

The Shannon capacity this channel is infinite in case there are no constraints on the channel inputs X_i; however, practical considerations always constrain the input in some manner. These input constraints are often defined in terms of the power of the input. For a channel input $(x_1, x_2, \ldots, x_n)$, the most common power constraints encountered are:

(**AP**): An average power constraint of $P > 0$, which says that

$$\sum_{i=1}^{n} x_i^2 \leq nP.$$

(**PP**): A peak power constraint of $A > 0$, which says that

$$|x_i| \leq A \quad \text{for all} \quad 1 \leq i \leq n.$$

(**APP**): An average and peak power constraint, consisting of (**AP**) and (**PP**) simultaneously.

The AWGN channel with the (**AP**) constraint was first analyzed by Shannon [33]. Shannon showed that the capacity C for this constraint is given by

$$C = \sup_{E[X^2] \leq P} I(X;Y) = \frac{1}{2} \log \left(1 + \frac{P}{\nu}\right), \tag{3.2}$$

and the supremum is attained when $X \sim \mathcal{N}(0, P)$. Here capacity is defined in the usual sense, due to Shannon. See Section 3.2 for a precise definition.

Compared to the (**AP**) constraint, fewer results exist about the (**PP**) constrained AWGN. The AWGN channel with the (**PP**) constraints was first analyzed by Smith [34]. Smith showed that the channel capacity C in this case is given by

$$C = \sup_{|X| \leq A} I(X;Y). \tag{3.3}$$

Unlike the (**AP**) case, the supremum in equation (3.3) does not have a closed form expression. Using tools from complex analysis, Smith established that the optimal input distribution attaining the supremum in equation (3.3) is discrete, and is supported on a finite number on points in the interval $[-A, A]$. He proposed an algorithm to numerically evaluate this optimal distribution, and thus the capacity. Smith also analyzed the (**APP**) constrained AWGN channel and derived similar results. In a related problem, Shamai & Bar-David [32] studied the quadrature Gaussian channel with (**APP**) constraints, and extended Smith's techniques to establish analogous capacity results for the same.

Our work is primarily concerned with a power constraint, which we call a (σ, ρ)-power constraint, defined as follows:

Definition. Let $\sigma, \rho \geq 0$. A codeword $(x_1, x_2, \ldots, x_n)$ is said to satisfy a (σ, ρ)-*power constraint* if

$$\sum_{j=k+1}^{l} x_j^2 \leq \sigma + (l-k)\rho \ , \ \forall \ 0 \leq k < l \leq n. \tag{3.4}$$

These constraints are motivated by *energy harvesting communication systems*, a research area which has seen a surge of interest in recent years. Energy harvesting (EH) is a process by which energy derived from an external source is captured, stored, and harnessed for applications. For example, harvested energy in the form of solar, thermal, or kinetic energy is converted into electrical energy using photoelectric, thermoelectric, or piezoelectric materials, and is used to power electronic devices. Energy which is harvested is generally present as ambient background and is free. EH devices are efficient, cheap, and require low maintenance, making them an attractive alternative to battery-powered devices. The problem of communicating over a noisy channel using harvested energy is encountered in a prominent application of EH: wireless sensor networks. Typically, sensor nodes used in such networks are battery-powered and thus have finite lifetimes. Since EH sensor nodes are capable of harvesting energy for their functioning, they have potentially infinite lifetimes and thereby have many advantages over their battery-powered counterparts [36].

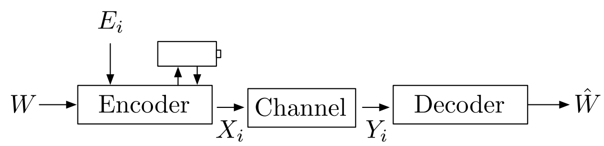

Figure 3.1: Block diagram of a general energy harvesting communication system

We can model communication scenarios like the "EH sensor node" via a general energy harvesting communication system shown in Figure 3.1. Here, the transmitter is capable of harvesting energy, and uses it to transmit a codeword X^n, corresponding to a message W. The transmitter has a battery to store the excess unutilized energy, which can be used for transmission later. The amount of energy harvested in time slot i, denoted by E_i, can be modeled as a stochastic process. The process E_i, along with the battery capacity, determines the power constraints that the codeword X^n has to satisfy. This codeword is transmitted over a noisy channel, and the receiver decodes W using the channel output Y^n. A natural channel to study in this setting is the classical additive Gaussian noise (AWGN) channel. Suppose we have a channel model as in Figure 3.2; namely, an AWGN channel with an energy harvesting transmitter which harvests a constant ρ amount of energy per time slot, and which has a battery of capacity σ attached to it.

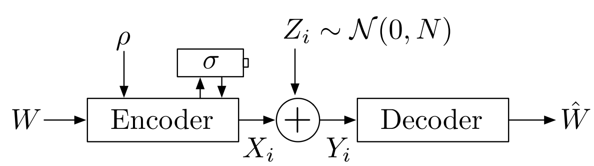

Figure 3.2: (σ, ρ)-power constrained AWGN channel

To understand the power constraints imposed on a transmitted codeword $(x_1, x_2, \ldots, x_n)$ in this scenario, we define a *state* σ_i, for each $i \geq 0$ as

$$\sigma_0 = \sigma, \text{ and } \sigma_{i+1} = \min(\sigma, \sigma_i + \rho - x_i^2) \ . \tag{3.5}$$

From the energy harvesting viewpoint, we can think of the state σ_i as the charge in the battery at time i before transmitting x_i, assuming the battery started out fully charged at time 0. Denote by $\mathcal{S}_n(\sigma, \rho) \subseteq \mathbb{R}^n$ the set

$$\mathcal{S}_n(\sigma, \rho) = \{x^n \in \mathbb{R}^n : \sigma_i \geq 0 \ , \ \forall \ 0 \leq i \leq n\}. \tag{3.6}$$

In words, the set $\mathcal{S}_n(\sigma, \rho)$ consists of sequences $(x_1, x_2, \ldots, x_n)$ such that at no point during its transmission, is there a need to overdraw the battery. Thus, this set is precisely the set of all possible length n sequences which the transmitter is capable of transmitting. Telescoping the minimum in equation (3.5), we get that for all $i \geq 0$,

$$\sigma_{i+1} = \min\left(\sigma,\ \sigma + \rho - x_i^2,\ \cdots,\ \sigma + i\rho - \sum_{j=1}^{i} x_j^2\right). \tag{3.7}$$

Using the condition $\sigma_i \geq 0$ for all i, we obtain another characterization of $\mathcal{S}_n(\sigma, \rho)$:

$$\mathcal{S}_n(\sigma, \rho) = \{x^n \in \mathbb{R}^n : \sum_{j=k+1}^{l} x_j^2 \leq \sigma + (l-k)\rho\ ,\ \forall\ 0 \leq k < l \leq n\}, \tag{3.8}$$

which is exactly the (σ, ρ)-power constraint defined in equation (3.4). It is interesting to note that such (σ, ρ)-constraints were originally introduced by Cruz [8, 9] in connection with the study of packet-switched networks. We first look at the (σ, ρ)-power constraint for the extreme cases; namely, $\sigma = 0$ and $\sigma = \infty$.

No battery:

Suppose that the battery capacity σ is 0; i.e., unused energy in a time slot cannot be stored for future transmissions. We can easily check that for a transmitted codeword $(x_1, x_2, ..., x_n)$, the power constraints

$$x_i^2 \leq \rho, \text{ for every } 1 \leq i \leq n \tag{3.9}$$

are necessary and sufficient to satisfy the inequalities in (3.4). Thus, the case of $\sigma = 0$ is simply the (**PP**) constraint of $\sqrt{\rho}$.

Infinite battery:

Consider the case where the battery capacity is now infinite, so that any unused energy can be saved for future transmissions. We assume that the battery is initially empty, but we can equally well assume it to start with any finite amount of energy in this scenario. The constraints imposed on a transmitted codeword $(x_1, x_2, ..., x_n)$ are

$$\sum_{i=1}^{k} x_i^2 \leq k\rho, \text{ for every } 1 \leq i \leq n. \tag{3.10}$$

It was shown by Ozel & Ulukus [25] that the strategy of initially saving energy and then using a Gaussian codebook achieves capacity, which is $\frac{1}{2}\log\left(1 + \frac{\rho}{N}\right)$. In fact, [25] considers not just constant E_i, but a more general case of i.i.d. E_i.

Finite battery:

An examination of equations (3.4) and (3.5) reveals that the energy constraint on the $n+1$-th symbol x_{n+1}, depends on the entire history of symbols transmitted up to time n. This infinite memory makes the exact calculation of channel capacity under these constraints a difficult task. For some recent work on *discrete* channels with finite batteries, we refer the reader to Tutuncuoglu et. al. [38, 39] and Mao & Hassibi [22]. An alternative model of an AWGN channel with a finite battery was also considered by Dong et. al. [12], where the authors established approximate capacity results for the same.

In this chapter, we will primarily focus on getting bounds on the channel capacity of an AWGN channel with (σ, ρ)-power constraints. Our work can be broadly divided into two parts; the first part deals with getting a lower bound, and the second part with getting an upper bound. The approach for both these parts relies on analyzing the geometric properties of the sets $\mathcal{S}_n(\sigma, \rho)$.

3.1 Summary of results

In what follows, we briefly describe our results.

3.1.1 Lower bound on capacity

We obtain a lower bound on the channel capacity in terms of the volume of $\mathcal{S}_n(\sigma, \rho)$. More precisely, we define $v(\sigma, \rho)$ to be the *exponential growth rate of volume* of the family $\{\mathcal{S}_n(\sigma, \rho)\}$:

$$v(\sigma, \rho) := \lim_{n \to \infty} \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_n(\sigma, \rho)), \tag{3.11}$$

where the limit can be shown to exist by subadditivity. Our first result is Theorem 3.3.2 in Section 3.3, which contains a lower bound on the channel capacity:

Theorem 3.3.2. *The capacity C of an AWGN channel with a (σ, ρ)-power constraint and noise power ν satisfies*

$$\frac{1}{2} \log \left(1 + \frac{e^{2v(\sigma,\rho)}}{2\pi e \nu}\right) \leq C \leq \frac{1}{2} \log \left(1 + \frac{\rho}{\nu}\right). \tag{3.12}$$

Having obtained this lower bound on C, it is natural to study the dependence of $v(\sigma, \rho)$ on its arguments. Theorem 3.4.1 in Section 3.4 establishes the following:

Theorem 3.4.1. *For a fixed ρ, $v(\sigma, \rho)$ is a monotonically increasing, continuous, and concave function of σ over $[0, \infty)$, with its range being $[\log 2\sqrt{\rho}, \frac{1}{2} \log 2\pi e \rho)$.*

In Section 3.5, we describe a numerical method to find $v(\sigma, \rho)$ for any value of the pair (σ, ρ). This calculated value can be used to compare the lower and upper bounds in Theorem 3.3.2 for different values of σ for a fixed ρ. From the energy-harvesting perspective, this comparison indicates the benefit that a finite battery of capacity σ has on the channel capacity. With this we conclude the first part of the section.

3.1.2 Upper bound on capacity

The upper bound on capacity in (3.12) is not satisfactory as it does not depend on σ. Our approach to deriving an improved upper bound on capacity also involves a volume calculation. However, the improved upper bound is not in terms of the volume of $\mathcal{S}_n(\sigma, \rho)$, but in terms of the volume of the Minkowski sum of $\mathcal{S}_n(\sigma, \rho)$ and a "noise ball." Let $B_n(\sqrt{n\nu})$ be the Euclidean ball of radius $\sqrt{n\nu}$. The Minkowski sum of $\mathcal{S}_n(\sigma, \rho)$ and $B_n(\sqrt{n\nu})$ (also called the parallel body of $\mathcal{S}_n(\sigma, \rho)$ at a distance $\sqrt{n\nu}$), is defined by

$$\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n\nu}) = \{x^n + z^n \mid x^n \in \mathcal{S}_n(\sigma, \rho), z^n \in B_n(\sqrt{n\nu})\} . \tag{3.13}$$

In Section 3.6, we prove the following upper bound on capacity:

Theorem 3.6.1. *The capacity C of an AWGN channel with a (σ, ρ)-power constraint and noise power ν satisfies*

$$C \leq \lim_{\epsilon \to 0_+} \limsup_{n \to \infty} \frac{1}{n} \log \frac{Vol(\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n(\nu + \epsilon)}\ ))}{Vol(B_n(\sqrt{n\nu}))}. \tag{3.14}$$

This motivates us to define a function $\ell : [0, \infty) \to \mathbb{R}$, giving the growth rate of the volume of the parallel body as follows:

$$\ell(\nu) := \limsup_{n \to \infty} \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n\nu}\ )). \tag{3.15}$$

The upper bound can be restated as

$$C \leq \limsup_{\epsilon \to 0_+} \left[\ell(\nu + \epsilon) - \frac{1}{2} \log 2\pi e \nu \right]. \tag{3.16}$$

To study the properties of $\ell(\cdot)$, we use the following result from convex geometry called *Steiner's formula*:

Theorem 3.6.2. *Let $K_n \subset \mathbb{R}^n$ be a compact convex set and let $B_n \subset \mathbb{R}^n$ be the unit ball. Denote by $\mu_j(K_n)$ the j-th intrinsic volume K_n, and by ϵ_j the volume of B_j. Then for $t \geq 0$,*

$$Vol(K_n \oplus tB_n) = \sum_{j=0}^{n} \mu_{n-j}(K_n) \epsilon_j t^j. \tag{3.17}$$

Intrinsic volume are a fundamental part of convex and integral geometry. They describe the global characteristics of a set, including the volume, surface area, mean width, and the Euler characteristic. For more details, we refer the reader to Schneider [30] and section 14.2 of Schneider & Weil [31].

In Section 3.7, we focus on the $\sigma = 0$ case for two reasons. Firstly, intrinsic volumes are notoriously hard to compute for arbitrary convex bodies. But when $\sigma = 0$, the set $\mathcal{S}_n(\sigma,\rho)$ is simply the cube $[-\sqrt{\rho},\sqrt{\rho}]^n$. The intrinsic volumes of a cube are well known in a closed form, which permits an explicit evaluation of $\ell(\nu)$. In his paper, Smith [34] numerically evaluated and plotted the capacity of a (**PP**) constrained AWGN channel. Based on the plots, Smith noted that as $\nu \to 0$, the channel capacity seemed to satisfy

$$C = \log 2A - \frac{1}{2}\log 2\pi e\nu + o(1), \tag{3.18}$$

where the $o(1)$ terms goes to 0 as $\nu \to 0$. He gave an intuitive explanation for this phenomenon as follows: Let X be the amplitude-constrained input, let $Z \sim \mathcal{N}(0,\nu)$ be the noise, and let Y be the channel output. Then for a small noise power ν, $h(Y) \approx h(X)$, and

$$\begin{aligned}
C &= \sup_X I(X;Y) \\
&= \sup_X h(Y) - h(Y|X) \\
&\approx \sup_X h(X) - h(Y|X) \\
&= \log 2A - \frac{1}{2}\log 2\pi e\nu.
\end{aligned}$$

Note that the crux of this argument is that when the noise power is small, $\sup_X h(Y) \approx \sup_X h(X) = \log 2A$. This argument can be made rigorous by establishing

$$\lim_{\nu\to 0}\left[\sup_X h(X+Z)\right] - \log 2A = 0. \tag{3.19}$$

Recall that our upper bound on capacity is $C \le \limsup_{\epsilon\to 0_+}\left\{\ell(\nu+\epsilon) - \frac{1}{2}\log 2\pi e\nu\right\}$. Since $\ell(0) = \log 2A$, the continuity of ℓ at 0 would lead to asymptotic upper bound which agrees with Smith's intuition. The following theorems provide our main result for the case of $\sigma = 0$:

Theorem 3.7.1. *The function $\ell(\nu)$ is continuous on $[0,\infty)$. For $\nu > 0$, we can explicitly compute $\ell(\nu)$ via the expression*

$$\ell(\nu) = H(\theta^*) + (1-\theta^*)\log 2A + \frac{\theta^*}{2}\log\frac{2\pi e\nu}{\theta^*}, \tag{3.20}$$

where H is the binary entropy function, and $\theta^* \in (0,1)$ satisfies

$$\frac{(1-\theta^*)^2}{{\theta^*}^3} = \frac{2A^2}{\pi\nu}.$$

Theorem 3.7.6. *The capacity C of an AWGN channel with an amplitude constraint of A, and with noise power ν, satisfies the following:*

1. *When the noise power $\nu \to 0$, capacity C is given by*

$$C = \log 2A - \frac{1}{2}\log 2\pi e\nu + O(\nu^{\frac{1}{3}}).$$

2. *When the noise power $\nu \to \infty$, capacity C is given by*

$$C = \frac{\alpha^2}{2} - \frac{\alpha^4}{4} + \frac{\alpha^6}{6} - \frac{5\alpha^8}{24} + O(\alpha^{10}),$$

where $\alpha = A/\sqrt{\nu}$.

We also establish a general entropy upper bound, which does not require the noise Z to be Gaussian:

Theorem 3.7.7. *Let $A, \nu \geq 0$. Let X and Z be random variables satisfying $|X| \leq A$ a.s. and Var$(Z) \leq \nu$. Then*

$$h(X + Z) \leq \ell(\nu). \tag{3.21}$$

In Section 3.8 we turn to the case of $\sigma > 0$. Unlike the $\sigma = 0$ case, the intrinsic volumes of $\mathcal{S}_n(\sigma, \rho)$ are not known in a closed form. For $n \geq 1$, we let $\{\mu_n(0), \cdots, \mu_n(n)\}$ be the intrinsic volumes of $\mathcal{S}_n(\sigma, \rho)$. The sequence of intrinsic volumes $\{\mu_n(\cdot)\}_{n\geq 1}$ forms a *sub-convolutive* sequence (analyzed in Section 2.1). Convergence properties of such sequences can be effectively studied using large deviation techniques; in particular, the Gärtner-Ellis theorem [10]. These convergence results for intrinsic volumes can be used in conjunction with Steiner's formula to establish results about ℓ and the asymptotic capacity of a (σ, ρ)-constrained channel in the low noise regime. Our main results here are:

Theorem 3.8.1. *Define $\ell(\nu)$ as*

$$\ell(\nu) = \limsup_{n\to\infty} \frac{1}{n} \log \mathit{Vol}(\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n\nu}\ )). \tag{3.22}$$

For $n \geq 1$, define $G_n : \mathbb{R} \to \mathbb{R}$ and $g_n : \mathbb{R} \to \mathbb{R}$ as

$$G_n(t) = \log \sum_{j=0}^{n} \mu_n(j) e^{jt}, \quad \textit{and} \quad g_n(t) = \frac{G_n(t)}{n}. \tag{3.23}$$

Define Λ to be the pointwise limit of the sequence of functions $\{g_n\}$, which we show exists. Let Λ^* be the convex conjugate of Λ. Then the following hold:

1. *$\ell(\nu)$ is continuous on $[0, \infty)$.*

2. *For $\nu > 0$,*
$$\ell(\nu) = \sup_{\theta \in [0,1]} \left[-\Lambda^*(1-\theta) + \frac{\theta}{2} \log \frac{2\pi e \nu}{\theta} \right]. \tag{3.24}$$

Theorem 3.8.10. *The capacity C of an AWGN channel with (σ, ρ)-power constraints and noise power ν satisfies the following:*

1. *When the noise power $\nu \to 0$, capacity C is given by*
$$C = v(\sigma, \rho) - \frac{1}{2} \log 2\pi e \nu + \epsilon(\nu),$$
where $\epsilon(\cdot)$ is a function such that $\lim_{\nu \to 0} \epsilon(\nu) = 0$.

2. *When noise power $\nu \to \infty$, capacity C is given by*
$$C = \frac{1}{2}\left(\frac{\rho}{\nu}\right) - \frac{1}{4}\left(\frac{\rho}{\nu}\right)^2 + \frac{1}{6}\left(\frac{\rho}{\nu}\right)^3 + O\left(\left(\frac{\rho}{\nu}\right)^4\right).$$

In Section 3.9, we describe a general framework which can be used to analyze power-constrained Gaussian channels. We analyze two types of power constraints within this framework. The first type of constraint is what we call a "block constraint", which is essentially a vector generalization of the amplitude constraint. The second type of constraint is the "super-convolutive constraint", a natural constraint to encounter which includes the average power constraint as a special case. In both cases, we establish capacity bounds and asymptotic capacity results just as in Sections 3.7 and 3.8.

3.2 Channel Capacity

We define channel capacity as per the usual convention [7]:

Definition 1. A $(2^{nR}, n)$ code for the AWGN channel with a (σ, ρ)-power constraint consists of the following:

1. A set of messages $\{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$

2. An encoding function $f : \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \to \mathcal{S}_n(\sigma, \rho)$, yielding codewords $f(1), \ldots, f(2^{\lfloor nR \rfloor})$

3. A decoding function $g : \mathbb{R}^n \to \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$

A rate R is said to be achievable if there exists a sequence of $(2^{nR}, n)$ codes such the that probability of decoding error diminishes to 0 as $n \to \infty$. The capacity of this channel is the supremum of all achievable rates.

Shannon's formula for channel capacity

$$C = \sup_{X} I(X; Y), \tag{3.25}$$

is valid if the channel is *memoryless*. For a channel with memory, one can often generalize this expression to

$$C = \lim_{n\to\infty} \left[\sup_{X^n} \frac{1}{n} I(X^n; Y^n) \right], \tag{3.26}$$

but this formula does not always hold. Dobrushin [11] showed that channel capacity is given by formula (3.26) for a class of channels called *information stable* channels. Checking information stability for specific channels can be quite challenging. Fortunately, in the case of a (σ, ρ)-power constrained AWGN channel, we can establish formula (3.26) without having to check for information stability. We prove the following theorem:

Theorem 3.2.1. *For $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the set of all probability distributions supported on $\mathcal{S}_n(\sigma, \rho)$. The capacity C of a (σ, ρ)-power constrained scalar AWGN channel is given by*

$$C = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}(x^n) \in \mathcal{F}_n} I(X^n; Y^n). \tag{3.27}$$

Proof. Let N be a positive integer. Without loss of generality, we can assume that coding is done for block lengths which are multiples N, say nN. For codes over such blocks, we relax the (σ, ρ) constraints as follows. For every transmitted codeword $(x_1, x_2, \cdots, x_{nN})$, each consecutive block of N symbols has to lie in $\mathcal{S}_N(\sigma, \rho)$; i.e.,

$$(x_{kN+1}, x_{kN+2}, ..., x_{(k+1)N}) \in \mathcal{S}_N(\sigma, \rho), \text{ for } 0 \le k \le n-1. \tag{3.28}$$

Note that this is indeed a relaxation because a codeword satisfying the constraint (3.28) is not guaranteed to satisfy the (σ, ρ)-constraints but any codeword satisfying the (σ, ρ)-constraints necessarily satisfies the constraint (3.28). The capacity of this channel C_N can be written as

$$C_N = \sup_{p_{X^N}(x^N) \in \mathcal{F}_N} I(X^N; Y^N). \tag{3.29}$$

This capacity provides an upper bound to NC for any choice of N. Thus, we have the bound

$$C \le \inf_{N} \frac{C_N}{N}. \tag{3.30}$$

To show that $\inf_N C_N/N$ is $\lim_N C_N/N$, we first note that

$$I(X_1^{M+N}; Y_1^{M+N}) \leq I(X_1^M; Y_1^M) + I(X_{M+1}^{M+N}; Y_{M+1}^{M+N}) \ .$$

Taking the supremum on both sides with $p_{X^{M+N}}$ ranging over $\mathcal{F}_{M+N}$,

$$C_{M+N} \leq \sup_{p_{X^{M+N}}(x^{M+N}) \in \mathcal{F}_{M+N}} \left(I(X_1^M; Y_1^M) + I(X_{M+1}^{M+N}; Y_{M+1}^{M+N}) \right) \tag{3.31}$$

$$\stackrel{(a)}{\leq} \sup_{p_{X^M}(x^M) \in \mathcal{F}_M} I(X_1^M; Y_1^M) + \sup_{p_{X^N}(x^N) \in \mathcal{F}_N} I(X_1^N; Y_1^N) \tag{3.32}$$

$$= C_M + C_N. \tag{3.33}$$

Here (a) follows due to the containment $\mathcal{F}_{M+N} \subseteq \mathcal{F}_M \times \mathcal{F}_N$. This calculation shows that $\{C_N\}$ is a sub-additive sequence. Applying Fekete's lemma [35] we conclude that $\lim_N C_N/N$ exists and equals $\inf_N C_N/N$, and thereby establish the upper bound

$$C \leq \lim_{N \to \infty} \frac{C_N}{N}. \tag{3.34}$$

We now show that C is lower bounded by $\lim_N C_N/N$. Given any $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \in \mathcal{S}_N$, the concatenated sequence $\mathbf{x}_1 \cdots \mathbf{x}_n$ need not always satisfy the (σ, ρ) power constraints. However, if we append $k = \lceil \frac{\sigma}{\rho} \rceil$ zeros to each $\mathbf{x}_i$ and then concatenate them, the $n(N+k)$ length string so formed lies in $\mathcal{S}_{n(N+k)}$. This is because transmitting $\lceil \frac{\sigma}{\rho} \rceil$ zeros after each $\mathbf{x}_i$ ensures that the state, as defined in equation (3.5), returns to σ before the transmission of $\mathbf{x}_{i+1}$ begins. Let us define a new set

$$\hat{\mathcal{S}}_N = \{x^{N+k} : x_1^N \in \mathcal{S}_N, x_{N+1}^{N+k} = \mathbf{0}\}.$$

The earlier discussion implies that

$$\underbrace{\hat{\mathcal{S}}_N \times \cdots \hat{\mathcal{S}}_N}_{n \text{ times}} \subseteq \mathcal{S}_{n(N+k)}. \tag{3.35}$$

Equation (3.35) implies that any block coding scheme which uses symbols from $\hat{\mathcal{S}}_N$ is also a valid coding scheme under the (σ, ρ) power constraints. The achievable rate for such a scheme can therefore provide a lower bound to C. This achievable rate is simply C_N, as the final k transmissions in each symbol carry no information. Thus the per transmission achievable rate is $\frac{C_N}{N+k}$, and we get that

$$C \geq \frac{C_N}{N+k}, \tag{3.36}$$

for all N. Taking the limit as $N \to \infty$, we arrive at the bound

$$C \geq \lim_{N \to \infty} \frac{C_N}{N}. \tag{3.37}$$

The containment (3.34), together with the inequality (3.37), completes the proof. $\square$

3.3 Lower-bounding capacity

Coding with the (σ, ρ) constraints can be thought of as trying to fit the largest number of centers of noise balls in $\mathcal{S}_n$, such that the noise balls are asymptotically approximately disjoint. One might therefore hope to get a packing based upper bound on capacity through the volume of $\mathcal{S}_n$. We shall show that the volume of $\mathcal{S}_n$ surprisingly yields a neat lower bound on capacity.

Let $V_n(\sigma, \rho)$ denote the volume of $\mathcal{S}_n(\sigma, \rho)$. We look at the exponential growth rate of this volume defined by

$$v(\sigma, \rho) := \lim_{n\to\infty} \frac{\log V_n(\sigma, \rho)}{n}. \tag{3.38}$$

Our first lemma is to establish the existence of the limit in the definition of $v(\sigma, \rho)$.

Lemma 3.3.1. $\lim_{n\to\infty} \frac{\log V_n(\sigma,\rho)}{n}$ *exists.*

Proof. The containment $\mathcal{S}_{m+n}(\sigma, \rho) \subseteq \mathcal{S}_m(\sigma, \rho) \times \mathcal{S}_n(\sigma, \rho)$ gives

$$V_{m+n}(\sigma, \rho) \leq V_m(\sigma, \rho) V_n(\sigma, \rho),$$

which implies

$$\log V_{m+n}(\sigma, \rho) \leq \log V_m(\sigma, \rho) + \log V_n(\sigma, \rho).$$

This shows that $\log V_n(\sigma, \rho)$ is a sub-additive sequence, and by Fekete's Lemma, the limit $\lim_{n\to\infty} \frac{\log V_n(\sigma,\rho)}{n}$ exists and is equal to $\inf_n \frac{\log V_n(\sigma,\rho)}{n}$ (which may a priori be $-\infty$). $\square$

Theorem 3.3.2. *The capacity C of an AWGN channel with (σ, ρ)-power constraints and noise power ν satisfies*

$$\frac{1}{2}\log\left(1 + \frac{e^{2v(\sigma,\rho)}}{2\pi e \nu}\right) \leq C \leq \frac{1}{2}\log\left(1 + \frac{\rho}{\nu}\right). \tag{3.39}$$

Proof. Clearly, C is upper bounded by the capacity for the $\sigma = \infty$ case (with zero initial battery condition), which by [25] is $\frac{1}{2}\log(1 + \frac{\rho}{\nu})$.

Let the noise $Z \sim \mathcal{N}(0, \nu)$. To prove the lower bound, recall the capacity expression in Theorem 3.2.1:

$$\begin{aligned} C &= \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} I(X^n; Y^n) && (3.40)\\ &= \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} h(Y^n) - h(Z^n) && (3.41)\\ &= \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} h(Y^n) - \frac{1}{2}\log 2\pi e \nu && (3.42) \end{aligned}$$

Thus, calculating capacity requires maximizing the output differential entropy $h(Y^n)$. Using Shannon's entropy power inequality, we have

$$e^{\frac{2h(Y^n)}{n}} \geq e^{\frac{2h(X^n)}{n}} + e^{\frac{2h(Z^n)}{n}}. \tag{3.43}$$

Thus,

$$\begin{aligned}
\sup_{p_{X^n}(x^n)\in\mathcal{F}_n} e^{\frac{2h(Y^n)}{n}} &\geq \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} e^{2\frac{h(X^n)}{n}} + 2\pi e\nu \\
&= e^{2\frac{\log V_n}{n}} + 2\pi e\nu.
\end{aligned}$$

Taking logarithms on both sides and letting n tend to infinity, we have

$$\lim_{n\to\infty} \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} \frac{h(Y^n)}{n} \geq \frac{1}{2}\log\left(e^{2v(\sigma,\rho)} + 2\pi e\nu\right), \tag{3.44}$$

which, combined with equation (3.42) concludes the proof. □

3.4 Properties of $v(\sigma,\rho)$

We can readily see that $v(\sigma,\rho)$ is monotonically increasing in both of its arguments. With a little more effort, we can also establish the following simple bounds for $v(\sigma,\rho)$:

$$\log 2\sqrt{\rho} \leq v(\sigma,\rho) \leq \log\sqrt{2\pi e\rho}. \tag{3.45}$$

To show the lower bound from inequality (3.45), observe that if x^n is such that for every $1 \leq i \leq n$,

$$|x_i| \leq \sqrt{\rho},$$

then the (σ,ρ)-constraints are satisfied. Thus, the cube $[-\sqrt{\rho},\sqrt{\rho}]^n$ of volume $(2\sqrt{\rho})^n$ lies inside the set $\mathcal{S}_n(\sigma,\rho)$, giving the lower bound

$$v(\sigma,\rho) \geq \log 2\sqrt{\rho}.$$

For the upper bound, we use the "total power" constraint,

$$x_1^2 + x_2^2 + \ldots + x_n^n \leq \sigma + n\rho,$$

which implies that $\mathcal{S}_n(\sigma,\rho) \subseteq B_n(\sqrt{\sigma+n\rho})$, where $B_n(\sqrt{\sigma+n\rho})$ is the Euclidean ball of radius $\sqrt{\sigma+n\rho}$. The volume $V_n(\sigma,\rho)$ of the set $\mathcal{S}_n(\sigma,\rho)$ is bounded above by the volume of $B_n(\sqrt{\sigma+n\rho})$, which gives

$$\begin{aligned}
v(\sigma,\rho) &\leq \lim_{n\to\infty}\frac{1}{n}\log\left(\frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}(\sigma+n\rho)^{\frac{n}{2}}\right) \\
&= \lim_{n\to\infty}\frac{1}{2}\left(\log\pi + \log(\sigma+n\rho) - \log\left(\frac{n}{2e}\right)\right) \\
&= \frac{1}{2}\log 2\pi e\rho.
\end{aligned}$$

Note that when $\rho = 0$, then $v(\sigma, 0) = -\infty$ for any value of σ. Henceforth, we assume $\rho > 0$. When $\sigma = 0$, the set $\mathcal{S}_n(\sigma, \rho)$ degenerates to the cube $[-\sqrt{\rho}, \sqrt{\rho}]^n$, which has the volume growth rate exponent of $\log 2\sqrt{\rho}$. It is clear that when $\sigma > 0$, the set $\mathcal{S}_n(\sigma, \rho)$ contains the cube $[-\sqrt{\rho}, \sqrt{\rho}]^n$, implying that

$$V_n(\sigma, \rho) > (2\sqrt{\rho})^n.$$

However, this does not immediately imply that $v(\sigma, \rho) > \log 2\sqrt{\rho}$. The following theorem is the main result of this section, where we show that such a strict inequality holds, and also prove some other properties of the function $v(\sigma, \rho)$:

Theorem 3.4.1. *For a fixed ρ, $v(\sigma, \rho)$ is a monotonically increasing, continuous, and concave function of $\sigma \in [0, \infty)$, with its range being $[\log 2\sqrt{\rho}, \frac{1}{2}\log 2\pi e \rho)$.*

Proof of Theorem 3.4.1. Theorem 3.4.1 relies on several lemmas. We state the lemmas here and defer their proofs to Appendix C.1. We first show that it is enough to prove the theorem for $\rho = 1$:

Lemma 3.4.2 (Proof in Appendix C.1.1)**.** *Let $v_1(\sigma) = v(\sigma, 1)$. Then $v(\sigma, \rho)$ depends on $v_1(\sigma/\rho)$ according to*

$$v(\sigma, \rho) = \log \sqrt{\rho} + v_1(\sigma/\rho). \tag{3.46}$$

Thus, a different value of ρ leads to a function $v(\sigma, \rho)$ which is essentially $v_1(\sigma)$ shifted by a constant. Therefore, if $v_1(\sigma)$ is monotonically increasing, continuous, and concave, so is $v(\sigma, \rho)$ for any other value of $\rho > 0$. In Lemmas 3.4.3 and 3.4.4, we establish that $v_1(\sigma)$ is a continuous and concave function on $[0, \infty)$:

Lemma 3.4.3 (Proof in Appendix C.1.2)**.** *The function $v_1(\sigma)$ is continuous on $[0, \infty)$.*

Lemma 3.4.4 (Proof in Appendix C.1.3)**.** *The function $v_1(\sigma)$ is concave on $[0, \infty)$.*

To finish the proof, we need to show that the limiting value of $v_1(\sigma)$ as $\sigma \to \infty$ is $\frac{1}{2}\log 2\pi e$. It is useful to define a quantity, which we call *burstiness of a sequence*, as follows: Let $\mathcal{A}_n$ denote the the n-dimensional ball of radius $\sqrt{n}$; i.e.,

$$\mathcal{A}_n := \left\{ x^n \;:\; \sum_{i=1}^{n} x_i^2 \le n \right\}.$$

Fix $x^n \in \mathcal{A}_n$. We associate a *burstiness* to each such sequence, defined by

$$\sigma(x^n) := \max_{0 \le k < l \le n} \left(\sum_{i=k+1}^{l} x_i^2 - (l-k) \right). \tag{3.47}$$

Let

$$\mathcal{A}_n(\sigma) = \{x^n \in \mathcal{A}_n \;:\; \sigma(x^n) \le \sigma\}.$$

Notice that $\mathcal{A}_n(\sigma) \subseteq \mathcal{S}_n(\sigma, 1)$. We have $\mathcal{A}_n(0) = [-1, 1]^n$ and $\mathcal{A}_n(n-1) = \mathcal{A}_n$. As σ increases from 0 to $n-1$, $\mathcal{A}_n(\sigma)$ increases from the cube to the entire sphere. We have the following lemma:

Lemma 3.4.5 (Proof in Appendix C.1.4)**.** *If there exists a sequence $\sigma(n)$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathit{Vol}(\mathcal{A}_n(\sigma(n))) = \frac{1}{2} \log 2\pi e \ , \quad \text{and} \tag{3.48a}$$

$$\lim_{n\to\infty} \frac{\sigma(n)}{n} = 0, \tag{3.48b}$$

then $\lim_{\sigma\to\infty} v_1(\sigma) = \frac{1}{2} \log 2\pi e$.

Note that the natural choice which satisfies condition (3.48a) is $\sigma(n) = n - 1$, but this does not satisfy condition (3.48b). To complete the proof, we show that $\sigma(n) = c\sqrt{n}$ for a suitable constant c satisfies both conditions of Lemma 3.4.5, and establish the following result:

Lemma 3.4.6 (Proof in Appendix C.1.5)**.** $\lim_{\sigma\to\infty} v(\sigma, 1) = \frac{1}{2} \log 2\pi e$.

This completes the proof of Theorem 3.4.1. □

3.5 Numerical method to compute $v(\sigma, \rho)$

In this section, we briefly discuss the numerical evaluation of $v(\sigma, \rho)$. This discussion is nontechnical and for all the technical details justifying the numerical method, we refer the reader to Appendix C.2.

Numerical computation of $v(\sigma, \rho)$ is enabled by exploiting the idea of state as defined in equation (3.5). However for ease of analysis and implementation, we define the state slightly differently. Given $(x_1, \ldots, x_n) \in \mathcal{S}_n(\sigma, 1)$, define

$$\phi_n = \begin{cases} \sigma_n & \text{if } \sigma_n < \sigma, \\ \sigma_{n-1} + 1 - x_n^2 & \text{if } \sigma_n = \sigma. \end{cases} \tag{3.49}$$

The state ϕ_n is a sum of two terms: σ_n, which is the amount of charge in the battery at time n, and the amount of energy wasted at time n due to the limited battery capacity. Note that energy is wasted only when $\sigma_n = \sigma$; i.e., when the battery becomes full. Setting $\phi_0 = \sigma$, equation (3.49) can also be written as

$$\phi_n = \begin{cases} \phi_{n-1} + 1 - x_n^2 & \text{if } \phi_{n-1} < \sigma, \\ \sigma + 1 - x_n^2 & \text{if } \phi_{n-1} \geq \sigma. \end{cases} \tag{3.50}$$

Consider the function $\Phi_n : \mathcal{S}_n(\sigma, 1) \to \mathbb{R}$, defined by $\Phi_n(x_1, \ldots, x_n) = \phi_n$. Thus, Φ_n maps a point in $\mathcal{S}_n(\sigma, 1)$ to its state at time n, as defined in equations (3.49) and (3.50). Let λ_n be the Lebesgue measure restricted to $\mathcal{S}_n(\sigma, 1)$. The function Φ_n induces a measure on $\mathbb{R}$, which we call ν_n. As $0 \leq \phi_n \leq \sigma + 1$ for all $x^n \in \mathcal{S}_n(\sigma, 1)$, we see that the measure ν_n is supported on $[0, \sigma + 1]$, giving

$$\nu_n([0, \sigma + 1]) = \mathrm{Vol}(\mathcal{S}_n(\sigma, 1)).$$

Suppose ν_n is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$; this implies existence of a density f_n corresponding to ν_n, which satisfies

$$\mathrm{Vol}(\mathcal{S}_n(\sigma, 1)) = \int_{x=0}^{\sigma+1} f_n(x) dx. \tag{3.51}$$

Given the state $\phi_n < \sigma$, the symbol x_{n+1} is constrained to lie in $[-\sqrt{\phi_n + 1}, \sqrt{\phi_n + 1}]$. Furthermore, given ϕ_n, the symbol x_{n+1} has the Lebesgue measure restricted to this set. Similarly, for $\phi_n \geq \sigma$, the conditional measure of x_{n+1} is the Lebesgue measure restricted to $[-\sqrt{\sigma + 1}, \sqrt{\sigma + 1}]$. Using equation (3.50), we can find a relation between the measures ν_n and ν_{n+1} as follows: For $\phi \in [0, \sigma + 1]$,

$$\begin{aligned}
F_{n+1}(\phi) &:= \nu_{n+1}((-\infty, \phi]) && (3.52)\\
&= \int_{x=0}^{\phi-1} \int_{t=0}^{x+1} \frac{f_n(x)}{\sqrt{t}} dt dx \\
&+ \int_{x=\phi-1}^{\sigma} \int_{t=x-(\phi-1)}^{x+1} \frac{f_n(x)}{\sqrt{t}} dt dx \\
&+ \int_{x=\sigma}^{\sigma+1} \int_{t=\sigma-(\phi-1)}^{\sigma+1} \frac{f_n(x)}{\sqrt{t}} dt dx && (3.53)\\
&= \int_{x=0}^{\phi-1} 2 f_n(x)[\sqrt{x+1}] dx \\
&+ \int_{x=\phi-1}^{\sigma} 2 f_n(x)[\sqrt{x+1} - \sqrt{x - (\phi - 1)}] dx \\
&+ \int_{x=\sigma}^{\sigma+1} 2 f_n(x)[\sqrt{\sigma+1} - \sqrt{\sigma - (\phi - 1)}] dx. && (3.54)
\end{aligned}$$

Differentiating F_{n+1}, we obtain

$$f_{n+1}(\phi) = \int_{\phi-1}^{\sigma} \frac{f_n(x)}{\sqrt{x - (\phi - 1)}} dx + \int_{\sigma}^{\sigma+1} \frac{f_n(x)}{\sqrt{\sigma - (\phi - 1)}} dx. \tag{3.55}$$

Define the integral operator A as follows:

$$A(x, t) = \begin{cases} \frac{1}{\sqrt{x+1-t}} & \text{if } 0 \leq x < \sigma \text{ and } 0 \leq t \leq x + 1, \\ \frac{1}{\sqrt{\sigma+1-t}} & \text{if } \sigma \leq x \leq \sigma + 1 \text{ and } 0 \leq t \leq \sigma + 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.56}$$

We can express equation (3.55) in another form,

$$f_{n+1}(t) = \int A(x,t) f_n(x) dx, \tag{3.57}$$

denoted by $f_{n+1} = A(f_n)$. Iterating this relation, we obtain

$$f_{n+1} = A^n f_1. \tag{3.58}$$

Our interest is in $v_1(\sigma)$, which by equations (3.51) and (3.58) is

$$v_1(\sigma) = \lim_{n\to\infty} \frac{1}{n} \log \int_0^{\sigma+1} A^{n-1} f_1(x) dx. \tag{3.59}$$

It seems natural to expect this limit to equal the largest eigenvalue of A. Our approach to finding the largest eigenvalue is to discretize A; let $h_n = \frac{\sigma+1}{n}$, and let A_n be an $(n+1)\times(n+1)$-matrix such that

$$A_n(i,j) = h_n \times A\big((i-1)h_n, (j-1)h_n\big) \quad \text{for} \quad 1 \le i,j \le n+1.$$

We can approximate the largest eigenvalue of the matrix A_n using standard methods and expect this value to tend to the largest eigenvalue of A as n becomes large.

Figure 3.3 shows the plot of $v_1(\sigma)$ obtained using the numerical procedure. Note that as σ becomes large, $v_1(\sigma)$ tends to the limit $\frac{1}{2}\log 2\pi e$ in a concave manner, as per Theorem 3.4.1.

We are now in a position to plot the bounds on capacity derived in Theorem 3.3.2. Figure 3.4 shows a plot of the lower and upper bounds for a fixed value of ρ $(=1)$ and for different values of the noise power ν. Note that even for relatively small values of σ, the volume based lower bound on capacity is close to the upper bound, which we recall is the channel capacity when $\sigma = \infty$. Thus, a small battery leads to significant gains in the capacity of a (σ,ρ)-power constrained AWGN channel.

3.6 Upper-bounding capacity

Theorem 3.3.2 states that $\frac{1}{2}\log\left(1+\frac{\rho}{\nu}\right)$ upper-bounds the channel capacity. This bound is not entirely satisfactory since it is independent of the value of σ. Furthermore, Figure 3.4 indicates that the lower bound and the upper bound do not converge asymptotically: as $\nu \to 0$, the lower bound is $v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu + O(\nu)$ and the upper bound is $\frac{1}{2}\log\frac{\rho}{\nu} + O(\nu)$, which differ by $O(1)$. This implies that either the upper bound, or the lower bound, or both, are loose in the low-noise regime. It is natural to expect the upper bound to be loose, since it disregards the effects of a finite σ on capacity. To obtain some insight on the low-noise capacity, it is useful to think of coding with the (σ,ρ)-constraints as trying to fit the largest number of centers of noise balls in $\mathcal{S}_n(\sigma,\rho)$,

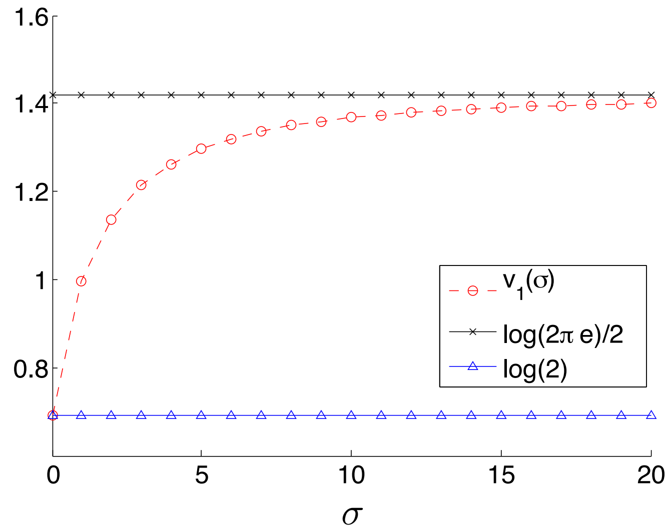

Figure 3.3: Graph of $v_1(\sigma)$ obtained numerically

such that the noise balls are asymptotically approximately disjoint. As the noise power ν decreases, so does the size of the noise balls, and one can imagine a very efficient packing of these small balls so that they occupy almost all the available space. The total number of balls one can pack is then roughly given by

$$\#\text{ of balls } \approx \frac{\text{Vol}(\mathcal{S}_n(\sigma,\rho))}{\text{Vol}(\text{Noise ball})}, \tag{3.60}$$

so the capacity is roughly

$$\begin{aligned} \frac{1}{n}\log \#\text{ of balls } &= \frac{1}{n}\log\frac{\text{Vol}(\mathcal{S}_n(\sigma,\rho)}{\text{Vol}(\text{Noise ball})} \qquad (3.61)\\ &\approx v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu. \qquad (3.62)\end{aligned}$$

We can make the statement in equation (3.62) rigorous, as follows:

Theorem 3.6.1. *Let $B_n(\sqrt{n\nu})$ be the n-dimensional Euclidean ball of radius $\sqrt{n\nu}$. The Minkowski sum of $\mathcal{S}_n(\sigma,\rho)$ and $B_n(\sqrt{n\nu})$ is the set*

$$\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu}) := \{x^n + z^n \mid x^n \in \mathcal{S}_n(\sigma,\rho), z^n \in B_n(\sqrt{n\nu})\}. \tag{3.63}$$

The capacity C of an AWGN channel with a (σ,ρ)-power constraint and noise power ν satisfies

$$C \le \lim_{\epsilon\to 0_+}\limsup_{n\to\infty}\frac{1}{n}\log \mathit{Vol}(\mathcal{S}_n(\sigma,\rho)\oplus B_n(\sqrt{n(\nu+\epsilon)}\,)) - \frac{1}{2}\log 2\pi e\nu. \tag{3.64}$$

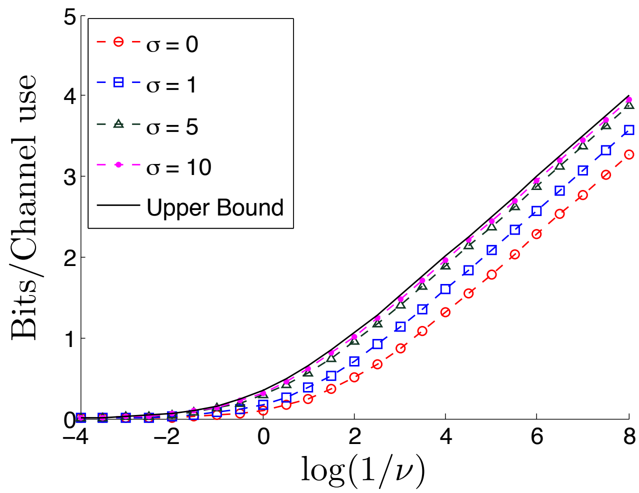

Figure 3.4: Capacity lower bounds for $\sigma = 0, 1, 5$, and 10, and the upper bound, from Theorem 3.3.2 plotted versus $\log(1/\nu)$

Proof. For $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the set of all probability distributions supported on $\mathcal{S}_n(\sigma,\rho)$. From Theorem 3.2.1, we know that the capacity C is given by

$$C = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}(x^n)\in\mathcal{F}_n} I(X^n;Y^n). \tag{3.65}$$

Let $p_{X^n}(x^n) \in \mathcal{F}_n$. Denote $\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n(\nu+\epsilon)}\ )$ by C_n, where $\epsilon > 0$, and let $\delta_n := \mathbb{P}\left(Y^n \notin C_n\right)$. By the law of large numbers, we have $\delta_n \to 0$. Let χ be the indicator variable for the event $\{Y^n \in C_n\}$. Then

$$\begin{aligned} h(Y^n) &= H(\delta_n) + \bar{\delta}_n h(Y^n|\chi = 1) + \delta_n h(Y^n|\chi = 0) \\ &\le H(\delta_n) + \bar{\delta}_n \log \mathrm{Vol}(C_n) + \delta_n h(Y^n|\chi = 0). \end{aligned} \tag{3.66}$$

where $\bar{a} = 1 - a$. Since $\|X^n\|^2 \le \sigma + n\rho$ with probability 1, we have the following bound on the power of Y^n:

$$E[\|Y^n\|^2] = E[\|X^n\|^2] + E[\|Z^n\|^2] \le \sigma + n\rho + n\nu.$$

This translates to the bound

$$E[\|Y^n\|^2 \mid \chi = 0] \le \frac{n(\rho+\nu+\sigma/n)}{\delta_n},$$

so

$$h(Y^n \mid \bar{\chi}) \le \frac{n}{2} \log \frac{2\pi e(\rho+\nu+\sigma/n)}{\delta_n}.$$

Substituting into inequality (3.66) and dividing by n gives

$$\frac{h(Y^n)}{n} \leq \frac{H(\delta_n)}{n} + \bar{\delta}_n \frac{\log \mathrm{Vol}(C_n)}{n} + \frac{\delta_n}{2} \log \frac{2\pi e(\rho + \nu + \sigma/n)}{\delta_n}.$$

Since this holds for any choice of $p_{X^n} \in \mathcal{F}_n$, we obtain

$$\sup_{p_{X^n} \in \mathcal{F}_n} \frac{1}{n} h(Y^n) \leq \frac{H(\delta_n)}{n} + \bar{\delta}_n \frac{\log \mathrm{Vol}(C_n)}{n} + \frac{\delta_n}{2} \log \frac{2\pi e(\rho + \nu + \sigma/n)}{\delta_n}.$$

Taking the limsup in n, we arrive at

$$\limsup_{n\to\infty} \sup_{p_{X^n} \in \mathcal{F}_n} \frac{1}{n} h(Y^n) \leq \limsup_{n\to\infty} \frac{\log \mathrm{Vol}(C_n)}{n} = \limsup_{n\to\infty} \frac{\log \mathrm{Vol}(\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n(\nu+\epsilon)}\ ))}{n}.$$

Taking the limit as $\epsilon \to 0_+$ and noting that the capacity is $\lim_{n\to\infty} \sup_{p_{X^n} \in \mathcal{F}_n} \frac{1}{n} h(Y^n) - \frac{1}{2} \log 2\pi e\nu$, we arrive at the bound in expression (3.64). $\square$

To simplify notation, define $\ell : [0, \infty) \to \mathbb{R}$ as

$$\ell(\nu) := \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu}\ )). \tag{3.67}$$

We can restate the upper bound in Theorem 3.6.1 as

$$C \leq \lim_{\epsilon\to 0_+} \ell(\nu+\epsilon) - \frac{1}{2} \log 2\pi e\nu. \tag{3.68}$$

If ℓ happens to be continuous at ν, we can drop the ϵ from inequality (3.68) to obtain a simplified expression

$$C \leq \ell(\nu) - \frac{1}{2} \log 2\pi e\nu. \tag{3.69}$$

Note that $\ell(0) = v(\sigma, \rho)$. The continuity of ℓ at $\nu = 0$ can be used to rigorously establish the asymptotic capacity expression in equation (3.62). These continuity properties will be established later in this section.

The upper bound expression involves the volume of the Minkowski sum of $\mathcal{S}_n(\sigma, \rho)$ with a ball. We state here a result from convex geometry called Steiner's formula [20], which gives an expression for the volume of such a Minkowski sum:

Theorem 3.6.2 (Steiner's formula)**.** *Let $K_n \subset \mathbb{R}^n$ be a compact convex set and let $B_n \subset \mathbb{R}^n$ be the unit ball. Denote by $\mu_j(K_n)$ the j-th intrinsic volume K_n, and by ϵ_j the volume of B_j. Then for $t \geq 0$,*

$$Vol(K_n \oplus tB_n) = \sum_{j=0}^{n} \mu_{n-j}(K_n) \epsilon_j t^j. \tag{3.70}$$

Steiner's formula states that the volume of $\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu})$ depends not only on the volumes of these sets, but also on the intrinsic volumes of $\mathcal{S}_n(\sigma,\rho)$. Intrinsic volumes are notoriously hard to compute even for simple enough sets such as polytopes [20]. So it is optimistic to expect a closed form expression for the intrinsic volumes of $\mathcal{S}_n(\sigma,\rho)$. Furthermore, the sets $\{\mathcal{S}_n(\sigma,\rho)\}$ evolve with the dimension n, and to compute the volume via Steiner's formula it is necessary to keep track of how the intrinsic volumes of these sets evolve with n.

As mentioned earlier, the case of $\sigma = 0$ is the amplitude-constrained Gaussian noise channel, the capacity of which was numerically evaluated by Smith [34]. In the following section, we concentrate on evaluating the upper bound for this special case.

3.7 The case of $\sigma = 0$

To simplify notation, we denote $A := \sqrt{\rho}$ in this section. We consider the scalar Gaussian noise channel with noise power ν and an input amplitude constraint of A. Let the capacity of this channel be C. Recall that the function $\ell(\nu)$ is defined as

$$\ell(\nu) = \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}([-A,A]^n \oplus B_n(\sqrt{n\nu})), \tag{3.71}$$

and the upper bound on channel capacity is given by

$$C \leq \lim_{\epsilon\to 0_+} \ell(\nu+\epsilon) - \frac{1}{2}\log 2\pi e\nu.$$

The main result of this section is as follows:

Theorem 3.7.1. *The function $\ell(\nu)$ is continuous on $[0,\infty)$. For $\nu > 0$, we can explicitly compute $\ell(\nu)$ via the expression*

$$\ell(\nu) = H(\theta^*) + (1-\theta^*)\log 2A + \frac{\theta^*}{2}\log\frac{2\pi e\nu}{\theta^*}, \tag{3.72}$$

where H is the binary entropy function, and $\theta^* \in (0,1)$ is the unique solution to

$$\frac{(1-\theta^*)^2}{{\theta^*}^3} = \frac{2A^2}{\pi\nu}.$$

Proof of Theorem 3.7.1. The proof of Theorem 3.7.1 relies on a number of lemmas. Here we shall merely state the lemmas and defer their proofs to Appendix C.3.

We first prove a lemma, which makes it possible to replace lim sup by lim in the expression of $\ell(\nu)$ given in equation (3.71).

Lemma 3.7.2 (Proof in Appendix C.3.1)**.** *For all $\nu \geq 0$, the limit*

$$\lim_{n\to\infty}\frac{1}{n}\log \mathit{Vol}([-A,A]^n \oplus B_n(\sqrt{n\nu}))$$

exists and is finite and equals $\ell(\nu)$, as defined in equation (3.71).

The special case of Steiner's formula (3.70) when K_n is the cube $[-A, A]^n$ and $t = \sqrt{n\nu}$ is given by

$$\mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})) = \sum_{j=0}^{n} \binom{n}{j} (2A)^{n-j} \epsilon_j (\sqrt{n\nu})^j, \tag{3.73}$$

where ϵ_j is the volume of the j-dimensional unit ball. Replacing ϵ_j in equation (3.73),

$$\begin{aligned}\mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})) &= \sum_{j=0}^{n} \binom{n}{j} (2A)^{n-j} \frac{\pi^{j/2}}{\Gamma(j/2+1)} (\sqrt{n\nu})^j &(3.74)\\ &= \sum_{j=0}^{n} \frac{\Gamma(n+1)}{\Gamma(n-j+1)\Gamma(j+1)} \frac{\pi^{j/2}}{\Gamma(j/2+1)} (2A)^{n-j} (\sqrt{n\nu})^j. &(3.75)\end{aligned}$$

Letting $\theta = \frac{j}{n}$, we rewrite the term inside the summation as

$$\frac{\Gamma(n+1)}{\Gamma(n(1-\theta)+1)\Gamma(n\theta+1)} \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)} (2A)^{n(1-\theta)} (\sqrt{n\nu})^{n\theta}. \tag{3.76}$$

For $\nu > 0$, define $f_n^\nu(\theta)$ as follows:

$$\begin{aligned}f_n^\nu(\theta) &= \frac{1}{n} \log \left(\frac{\Gamma(n+1)}{\Gamma(n(1-\theta)+1)\Gamma(n\theta+1)} \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)} (2A)^{n(1-\theta)} (\sqrt{n\nu})^{n\theta} \right) &(3.77)\\ &= \frac{1}{n} \log \left(\frac{\Gamma(n+1) n^{n\theta/2}}{\Gamma(n(1-\theta)+1)\Gamma(n\theta+1)\Gamma(n\theta/2+1)} \right)\\ &\quad + (1-\theta) \log 2A + \theta \log \sqrt{\nu} + \frac{\theta}{2} \log \pi. &(3.78)\end{aligned}$$

Note that $f_n^\nu(\theta)$ is defined for all $n \in \mathbb{N}$, for all $\theta \in [0, 1]$, and for all $\nu > 0$. Using this notation, we can rewrite the volume as

$$\mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})) = \sum_{j=0}^{n} e^{n f_n^\nu(j/n)}. \tag{3.79}$$

We argue that since the volume is a sum of $n+1$ terms, the exponential growth rate of the volume is determined by the growth rate of the largest term amongst these $n+1$ terms. To be precise, we define

$$\hat{\theta}_n = \arg\max_{j/n} f_n^\nu(j/n), \tag{3.80}$$

and prove the following lemma:

Lemma 3.7.3 (Proof in Appendix C.3.2)**.** *The limit* $\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n)$ *exists and equals* $\ell(\nu)$.

The next few lemmas aim to identify the limit of $f_n^\nu(\hat{\theta}_n)$. We first show that the functions $f_n^\nu(\cdot)$ converge uniformly to a limit function $f^\nu(\cdot)$.

Lemma 3.7.4 (Proof in Appendix C.3.3)**.** *The sequence of functions* $\{f_n^\nu\}_{n=1}^\infty$ *converges uniformly for all* $\theta \in [0, 1]$ *to a function* f^ν *given by*

$$f^\nu(\theta) = H(\theta) + (1-\theta)\log 2A + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}, \tag{3.81}$$

where $H(\theta) = -\theta\log\theta - (1-\theta)\log(1-\theta)$ *is the binary entropy function.*

With this uniform convergence in hand, we show that the limit of $f_n^\nu(\hat{\theta}_n)$ can be expressed as follows:

Lemma 3.7.5 (Proof in Appendix C.3.4)**.** *We claim that*

$$\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n) = \max_\theta f^\nu(\theta), \tag{3.82}$$

and therefore

$$\ell(\nu) = \max_\theta f^\nu(\theta). \tag{3.83}$$

We are now in a position to prove the continuity of $\ell(\nu)$. Fix a $\nu_0 > 0$, and let $\epsilon > 0$ be given. Choose a $\delta > 0$ such that for all $\nu \in (\nu_0 - \delta, \nu_0 + \delta)$,

$$||f^\nu - f^{\nu_0}||_\infty < \epsilon.$$

We can verify from equation (3.81) that picking such a δ is indeed possible. This implies

$$|\sup_\theta f^\nu(\theta) - \sup_\theta f^{\nu_0}(\theta)| < \epsilon. \tag{3.84}$$

Using Lemma 3.7.5, this implies

$$|\ell(\nu) - \ell(\nu_0)| < \epsilon, \tag{3.85}$$

which establishes continuity of ℓ at all points $\nu_0 > 0$.

To show continuity at 0, we first explicitly evaluate $\ell(\nu)$. Let $\theta^*(\nu) = \arg\max_\theta f^\nu(\theta)$. Using Lemma 3.7.5, we have $\ell(\nu) = f^\nu(\theta^*(\nu))$. Recall the expression for $f^\nu(\theta)$:

$$f^\nu(\theta) = H(\theta) + (1-\theta)\log 2A + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}. \tag{3.86}$$

Differentiating $f^{\nu}(\theta)$ with respect to θ,

$$\frac{d}{d\theta} f^{\nu}(\theta) = \log \frac{1-\theta}{\theta} + \log \sqrt{\nu} - \log 2A + \frac{1}{2} \log 2\pi e - \frac{\log e}{2} - \frac{1}{2} \log \theta. \tag{3.87}$$

Setting the derivative equal to 0 gives

$$\log \frac{1-\theta}{\theta} + \log \sqrt{\nu} - \log 2A + \frac{1}{2} \log 2\pi e - \frac{\log e}{2} - \frac{1}{2} \log \theta = 0. \tag{3.88}$$

Simplifying this and removing the logarithms, we arrive at

$$\frac{(1-\theta)^2}{\theta^3} = \frac{2A^2}{\pi\nu}. \tag{3.89}$$

The function $\frac{(1-\theta)^2}{\theta^3}$ tends to $+\infty$ as $\theta \to 0_+$, and equals 0 when $\theta = 1$. Thus, equation (3.89) has at least one solution in the interval $(0, 1)$. We can easily check that $\frac{(1-\theta)^2}{\theta^3}$ is strictly decreasing in $(0, 1)$, and thus this solution must be unique. The optimal $\theta^*(\nu)$ satisfies the cubic equation (3.89), and we can see that

$$\lim_{\nu \to 0} \theta^*(\nu) = 0. \tag{3.90}$$

Using equations (3.89) and (3.90), we have

$$\lim_{\nu \to 0} \frac{\nu}{\theta^*(\nu)^3} = \frac{2A^2}{\pi}. \tag{3.91}$$

Thus,

$$\begin{aligned}
\lim_{\nu \to 0} \ell(\nu) &= \lim_{\nu \to 0} H(\theta^*(\nu)) + (1 - \theta^*(\nu)) \log 2A + \frac{\theta^*(\nu)}{2} \log \frac{2\pi e \nu}{\theta^*(\nu)} && (3.92)\\
&\overset{(a)}{=} \log 2A + \lim_{\nu \to 0} \frac{\theta^*(\nu)}{2} \log \frac{2\pi e \nu}{\theta^*(\nu)} && (3.93)\\
&\overset{(b)}{=} \log 2A && (3.94)\\
&= \ell(0), && (3.95)
\end{aligned}$$

where in (a) we used equation (3.90), and in (b) we used equation (3.91). This shows that ℓ is continuous over $[0, \infty)$, and concludes the proof of Theorem 3.7.1. □

The above bound can also be used to prove an asymptotic capacity result. We prove the following theorem:

Theorem 3.7.6. *The capacity C of an AWGN channel with an amplitude constraint of A, and with noise power ν, satisfies the following:*

1. *When the noise power $\nu \to 0$, capacity C is given by*

$$C = \log 2A - \frac{1}{2}\log 2\pi e\nu + O(\nu^{\frac{1}{3}}).$$

2. *When the noise power $\nu \to \infty$, capacity C is given by*

$$C = \frac{\alpha^2}{2} - \frac{\alpha^4}{4} + \frac{\alpha^6}{6} - \frac{5\alpha^8}{24} + O(\alpha^{10}),$$

where $\alpha = A/\sqrt{\nu}$.

Proof of Theorem 3.7.6. Note that all the logarithms in this proof are assumed to be to base e.

1. Using the lower bound in Theorem 3.3.2,

$$\begin{aligned} C &\geq \frac{1}{2}\log\left(1 + \frac{(2A)^2}{2\pi e\nu}\right) && (3.96)\\ &= \log 2A - \frac{1}{2}\log 2\pi e\nu + \log\left(1 + \frac{2\pi e\nu}{(2A)^2}\right) && (3.97)\\ &= \log 2A - \frac{1}{2}\log 2\pi e\nu + O(\nu). && (3.98) \end{aligned}$$

For the upper bound, we have

$$\begin{aligned} \lim_{\nu\to 0}\ell(\nu) &= \log 2A + \lim_{\nu\to 0}\left[H(\theta^*) - \theta^*\log 2A + \frac{\theta^*}{2}\log\frac{2\pi e\nu}{\theta^*}\right] && (3.99)\\ &= \log 2A + \lim_{\nu\to 0} -(1-\theta^*)\log(1-\theta^*) + \frac{\theta^*}{2}\log\frac{\nu}{\theta^{*3}} + \frac{\theta^*}{2}\log\frac{\pi e}{2A^2}. && (3.100) \end{aligned}$$

Let $c = \left(\frac{\pi}{2A^2}\right)^{1/3}$. Using equation (3.91), we can check that as $\nu \to 0$,

$$\begin{aligned} -(1-\theta^*)\log(1-\theta^*) &= c\nu^{1/3} + o(\nu^{1/3}),\\ \frac{\theta^*}{2}\log\frac{\nu}{\theta^{*3}} &= \frac{-3c\log c}{2}\nu^{1/3} + o(\nu^{1/3}),\\ \frac{\theta^*}{2}\log\frac{\pi e}{2A^2} &= \left(\frac{c}{2} + \frac{3c\log c}{2}\right)\nu^{1/3} + o(\nu^{1/3}). \end{aligned}$$

This gives the following asymptotic upper bound as $\nu \to 0$:

$$C \leq \log 2A - \frac{1}{2}\log 2\pi e\nu + \frac{3c}{2}\nu^{1/3} + o(\nu^{1/3}). \qquad (3.101)$$

From equations (3.98) and (3.101), our claim follows.

2. As noted by Smith [34], for large ν the optimal input distribution is discrete, and is supported equally on the two points $-A$ and $+A$. The output Y is then distributed as

$$Y \sim p_Y(y) = \frac{1}{2}\exp\left(\frac{(y-A)^2}{2\nu}\right) + \frac{1}{2}\exp\left(\frac{(y+A)^2}{2\nu}\right) \tag{3.102}$$

Capacity is then given by

$$C = h(p_Y) - \frac{1}{2}\log 2\pi e\nu. \tag{3.103}$$

The entropy term $h(p_Y)$ can be manipulated as in [24] to arrive at

$$h(p_Y) = \frac{1}{2}\log 2\pi e\nu + \alpha^2 - \frac{2}{\sqrt{(2\pi)}\alpha}e^{-\alpha^2/2}\int_0^\infty e^{-y^2/2\alpha^2}\cosh(y)\ln\cosh(y)dy, \tag{3.104}$$

where $\alpha = A/\sqrt{\nu}$. Let

$$f(\alpha) = \frac{2}{\sqrt{2\pi}}\int_0^\infty e^{\frac{-y^2}{2\alpha^2}}\cosh(y)\ln\big(\cosh(y)\big)dy.$$

We consider the Taylor series expansion of $\cosh(y)\ln\big(\cosh(y)\big)$ at $y = 0$, and arrive at

$$\cosh(y)\log\big(\cosh(y)\big) = \frac{y^2}{2} + \frac{y^4}{6} + \frac{y^6}{720} + \frac{y^8}{630} + O\left(y^9\right). \tag{3.105}$$

Using the following definite integral expression,

$$\int_0^\infty e^{\frac{-y^2}{2\alpha^2}} y^{2k} dy = 2^{k-\frac{1}{2}}\left(\alpha^2\right)^{k+\frac{1}{2}}\Gamma\left(k+\frac{1}{2}\right), \tag{3.106}$$

and substituting, we obtain

$$f(\alpha) = \frac{\alpha^3}{2} + \frac{\alpha^5}{2} + \frac{\alpha^7}{48} + \frac{\alpha^9}{6} + O(\alpha^{11}). \tag{3.107}$$

Thus,

$$\begin{aligned} h(p_Y) &= \frac{1}{2}\log 2\pi e\nu + \alpha^2 \\ &\quad - \left(\frac{\alpha^2}{2} + \frac{\alpha^4}{2} + \frac{\alpha^6}{48} + \frac{\alpha^8}{6} + O(\alpha^{10})\right)\left(1 - \frac{\alpha^2}{2} + \frac{\alpha^4}{8} - \frac{\alpha^6}{48} + O(\alpha^8)\right) \qquad (3.108) \\ &= \frac{1}{2}\log 2\pi e\nu + \frac{\alpha^2}{2} - \frac{\alpha^4}{4} + \frac{\alpha^6}{6} - \frac{5\alpha^8}{24} + O(\alpha^{10}). \qquad (3.109) \end{aligned}$$

Capacity is therefore given by

$$C = \frac{\alpha^2}{2} - \frac{\alpha^4}{4} + \frac{\alpha^6}{6} - \frac{5\alpha^8}{24} + O(\alpha^{10}).$$

This establishes the claim. Shannon [33] had proved that capacity at high noise for the peak power constrained (by A^2) AWGN channels is essentially the same as that of an average power constrained (by A^2) AWGN; i.e.,

$$\begin{aligned} C \approx \frac{1}{2}\log(1 + A^2/\nu) &= \frac{1}{2}\log(1+\alpha^2) \\ &= \frac{\alpha^2}{2} - \frac{\alpha^4}{4} + \frac{\alpha^6}{6} - \frac{\alpha^8}{8} + O(\alpha^{10}). \end{aligned}$$

It is interesting to note that the first three terms of this approximation agrees with the actual capacity.

$\square$

We can use Theorem 3.7.1 to numerically evaluate $\theta^*(\nu)$ and plot the corresponding upper bound from Theorem 3.6.1. Figure 3.5 shows the resulting plot. Note that the upper bound from Theorem 3.3.2 is not asymptotically tight in the low-noise regime, but the new upper bound is asymptotically tight.

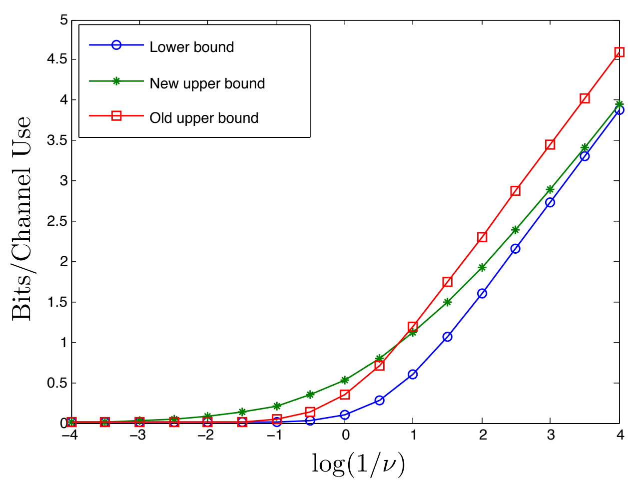

Figure 3.5: For the AWGN with an amplitude constraint of 1, the new upper bound and the lower bound converge asymptotically as $\nu \to 0$

In Theorem 3.7.1, we essentially carried out a volume computation which answered the question: How does the volume of the Minkowski sum of a cube and a ball grow? The upper bound on capacity is then a consequence of the following facts:

1. The channel capacity depends on the maximum output entropy $h(Y^n)$.
2. The random variable Y^n is (almost entirely) supported on the sum of a cube and a ball.
3. The entropy of Y^n is bounded from above by the logarithm of the volume of its (almost) support.

Intuitively, points 2 and 3 should not depend on Z being Gaussian, but only on Z^n being almost entirely supported on $B_n(\sqrt{n\nu})$. We make this intuition precise in the following theorem:

Theorem 3.7.7 (Proof in Appendix C.3.5)**.** *Let $A, \nu \geq 0$. Let X and Z be random variables satisfying $|X| \leq A$ a.s. and Var$(Z) \leq \nu$. Then*

$$h(X + Z) \leq \ell(\nu), \tag{3.110}$$

where $\ell(\nu)$ is as defined in equation (3.71).

By Theorem 3.7.7, we can assert that the capacity C of any channel with input amplitude constrained by A and with an additive noise Z with power at most ν is bounded from above according to

$$C = \sup_{|X| \leq A} I(X; X + Z) \leq \ell(\nu) - h(Z). \tag{3.111}$$

Noting that Var(Y) $\leq A^2 + \nu$, we also have the upper bound

$$C \leq \frac{1}{2} \log 2\pi e(\nu + A^2) - h(Z). \tag{3.112}$$

giving

$$C \leq \min \left(\ell(\nu) - h(Z), \frac{1}{2} \log 2\pi e(\nu + A^2) - h(Z) \right). \tag{3.113}$$

From Figure 3.5, it is interesting to note that there for large values of ν, the bound in inequality (3.112) is better, whereas for small values of ν, the bound in inequality (3.111) is better. Both of these bounds are asymptotically tight as $\nu \to \infty$, but only inequality (3.111) is tight for $\nu \to 0$.

3.8 The case of $\sigma > 0$

In this section, our aim is to parallel the upper-bounding technique used in Section 3.7 and obtain analogues of Theorem 3.7.1 and Theorem 3.7.6, when σ is strictly greater than 0. When $\sigma > 0$, the set $\mathcal{S}_n(\sigma, \rho)$ is no longer an easily identifiable set like the n-dimensional cube from Section 3.7. In particular, the intrinsic volumes of $\mathcal{S}_n(\sigma, \rho)$

do not have a closed form expression. Despite this difficulty, we shall see that it is still possible to obtain results similar to those in Section 3.7.

Our main result in this section is the following:

Theorem 3.8.1. *Define $\ell(\nu)$ as*

$$\ell(\nu) = \limsup_{n\to\infty} \frac{1}{n} \log Vol(\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n\nu}\,)). \tag{3.114}$$

For $n \geq 1$, denote the intrinsic volumes of $\mathcal{S}_n(\sigma, \rho)$ by $\mu_n(i)$ for $0 \leq i \leq n$ and define $G_n : \mathbb{R} \to \mathbb{R}$ and $g_n : \mathbb{R} \to \mathbb{R}$ as

$$G_n(t) = \log \sum_{j=0}^{n} \mu_n(j) e^{jt}, \quad g_n(t) = \frac{G_n(t)}{n}. \tag{3.115}$$

Define Λ to be the pointwise limit of the sequence of functions $\{g_n\}$, which we will show exists. Let Λ^* be the convex conjugate of Λ. Then the following hold:

1. *$\ell(\nu)$ is continuous on $[0, \infty)$.*
2. *For $\nu > 0$,*
$$\ell(\nu) = \sup_{\theta\in[0,1]} \left[-\Lambda^*(1-\theta) + \frac{\theta}{2} \log \frac{2\pi e\nu}{\theta} \right]. \tag{3.116}$$

Proof of Theorem 3.8.1. Note that for the statement of Theorem 3.8.1 to make sense, several results need to be established. We establish these in the Lemmas 3.8.2 and 3.8.3, where we prove the following:

Lemma 3.8.2 (Proof in Appendix C.4.1)**.** *For all $n \geq 1$, the set $\mathcal{S}_n(\sigma, \rho)$ is a convex set, and therefore it has well defined intrinsic volumes $\{\mu_n(i)\}_{i=0}^n$.*

Lemma 3.8.3 (Proof in Appendix C.4.2)**.** *The following results hold:*

1. *The functions $\{g_n\}$ converge pointwise to a function $\Lambda(t) : \mathbb{R} \to \mathbb{R}$ given by*
$$\Lambda(t) := \lim_{n\to\infty} g_n(t). \tag{3.117}$$
2. *The convex conjugate of Λ, denoted by Λ^*, has its domain the set $[0, 1]$.*

By Lemma 3.8.2, we can use Steiner's formula for the convex set $\mathcal{S}_n(\sigma, \rho)$ to get

$$\text{Vol}(\mathcal{S}_n(\sigma, \rho) \oplus B_n(\sqrt{n\nu}\,)) = \sum_{j=0}^{n} \mu_n(n-j)\epsilon_j \sqrt{n\nu}^{\,j}. \tag{3.118}$$

Define the functions $a_n(\theta)$ and $b_n^\nu(\theta)$ for $\theta \in [0,1]$ as follows. The function $a_n(\theta)$ is obtained by linearly interpolating the values of $a_n(j/n)$, where the value of $a_n(j/n)$ is given by:

$$a_n\left(\frac{j}{n}\right) = \frac{1}{n}\log \mu_n(n-j) \quad \text{for} \quad 0 \le j \le n. \tag{3.119}$$

The function $b_n^\nu(\theta)$ is given by

$$b_n^\nu(\theta) = \frac{1}{n}\log \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)}(n\nu)^{n\theta/2} \quad \text{for} \quad \theta \in [0,1]. \tag{3.120}$$

Define $f_n^\nu : [0,1] \to \mathbb{R}$ as

$$f_n^\nu(\theta) := f^\nu(n,\theta) = a_n(\theta) + b_n^\nu(\theta). \tag{3.121}$$

With this notation, we can rewrite equation (3.118) as

$$\mathrm{Vol}(\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu})) = \sum_{j=0}^{n} e^{nf^\nu(n,j/n)}. \tag{3.122}$$

Just as in the proof of Theorem 3.7.1, we want to establish the convergence of $f_n^\nu(\cdot)$ to some function $f^\nu(\cdot)$. Proving the convergence of $b_n^\nu(\cdot)$ is not hard, but proving the convergence of $a_n(\cdot)$ requires the application of Lemmas 3.8.4 and 3.8.5 given below. In Lemma 3.8.4 we establish the following:

Lemma 3.8.4 (Proof in Appendix C.4.3)**.** *For each n, the following holds:*

1. *The function $a_n(\cdot)$ is concave.*
2. *The function $b_n^\nu(\cdot)$ is concave.*
3. *The function $f_n^\nu(\cdot)$ is concave.*

In Lemma 3.8.5, we show that the intrinsic volumes of $\{\mathcal{S}_n(\sigma,\rho)\}$ satisfy a large deviations-type result, detailed below.

Lemma 3.8.5 (Proof in Appendix C.4.4)**.** *Define a sequence of measures supported on $[0,1]$ by*

$$\mu_{n/n}\left(\frac{j}{n}\right) := \mu_n(j) \quad \textit{for} \quad 0 \le j \le n. \tag{3.123}$$

The following bounds hold:

1. *Let $I \subseteq \mathbb{R}$ be a closed set. The family of measures $\{\mu_{n/n}\}$ satisfies the large deviation upper bound*

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_{n/n}(I) \le -\inf_{x\in I} \Lambda^*(x). \tag{3.124}$$

2. *Let $F \subseteq \mathbb{R}$ be an open set. The family of measures $\{\mu_{n/n}\}$ satisfies the large deviations lower bound*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) \geq -\inf_{x\in F} \Lambda^*(x). \tag{3.125}$$

Using the concavity and large deviations-type convergence from the two previous lemmas, we now prove the convergence of $\{f_n^\nu\}$ in the following lemma.

Lemma 3.8.6 (Proof in Appendix C.4.5)**.** *The following convergence results hold:*

1. *The sequence of functions $\{a_n\}$ converges uniformly to $-\Lambda^*(1-\theta)$ on $[0,1]$.*
2. *The sequence of functions $\{b_n^\nu\}$ converges uniformly to the function $\frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}$ on $[0,1]$.*
3. *The sequence of functions $\{f_n^\nu\}$ converges uniformly to a function f^ν on the interval $[0,1]$, where f^ν is given by*

$$f^\nu(\theta) = -\Lambda^*(1-\theta) + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}.$$

We are now in a position to express $\ell(\nu)$ in terms of the limit function f^ν. Let

$$\hat{\theta}_n = \arg\max_{j/n} f_n^\nu(j/n).$$

In Lemma 3.8.7 we prove the following:

Lemma 3.8.7 (Proof in Appendix C.4.6)**.** *The following equality holds:*

$$\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n) = \max_\theta f^\nu(\theta). \tag{3.126}$$

Lemma 3.8.8 (Proof in Appendix C.4.7)**.** *The following equality holds:*

$$\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n) = \ell(\nu), \tag{3.127}$$

and therefore

$$\ell(\nu) = \sup_\theta f^\nu(\theta). \tag{3.128}$$

Part 2 of Theorem 3.8.1 follows from Lemma 3.8.8. We now concentrate on proving the continuity of $\ell(\nu)$. We first show continuity at all points $\nu \neq 0$.

Let $\nu_0 > 0$, and let $\epsilon > 0$ be given. Choose a $\delta > 0$ such that for all $\nu \in (\nu_0-\delta, \nu_0+\delta)$,

$$||f^\nu - f^{\nu_0}||_\infty < \epsilon.$$

This implies

$$|\sup_\theta f^\nu(\theta) - \sup_\theta f^{\nu_0}(\theta)| < \epsilon \implies |\ell(\nu) - \ell(\nu_0)| < \epsilon, \tag{3.129}$$

which establishes continuity of ℓ at all points $\nu_0 > 0$.

Turning towards the $\nu = 0$ case, we define

$$\theta^*(\nu) = \arg\max_\theta f^\nu(\theta). \tag{3.130}$$

Proving the continuity of ℓ at $\nu = 0$ is slightly more challenging than the corresponding proof in Theorem 3.7.1 from Section 3.7, since we do not know $\theta^*(\nu)$ explicitly in terms of ν. Despite this, we can still prove the following lemma:

Lemma 3.8.9 (Proof in Appendix C.4.8)**.** *The following equality holds:*

$$\limsup_{\nu\to 0} \theta^*(\nu) = 0. \tag{3.131}$$

Now let $\nu_0 = 0$ and let $\epsilon > 0$ be given. Using continuity of Λ^*, choose an $\eta > 0$ such that

$$|-\Lambda^*(1-\theta) - v(\sigma,\rho)| < \epsilon/2 \text{ for all } \theta \in [0,\eta). \tag{3.132}$$

Using Lemma 3.8.9, choose a δ_1 such that

$$\theta^*(\nu) < \eta \text{ for all } \nu \in [0,\delta_1). \tag{3.133}$$

For all $\nu \in [0,\delta_1)$, we have

$$\begin{aligned}
\ell(\nu) &= \sup_\theta \left[-\Lambda^*(1-\theta) + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}\right] && (3.134)\\
&= -\Lambda^*(\theta^*(\nu)) + \frac{\theta^*(\nu)}{2}\log\frac{2\pi e\nu}{\theta^*(\nu)} && (3.135)\\
&\overset{(a)}{<} v(\sigma,\rho) + \frac{\epsilon}{2} + \sup_\theta \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta} && (3.136)\\
&\overset{(b)}{\le} v(\sigma,\rho) + \frac{\epsilon}{2} + \pi\nu && (3.137)
\end{aligned}$$

where (a) follows by inequalities (3.132) and (3.133), and (b) follows from an evaluation of the supremum in (a). Choose $\delta_2 = \frac{\epsilon}{2\pi}$, and choose $\delta = \min(\delta_1,\delta_2)$. We now have that for all $\nu \in [0,\delta)$,

$$\ell(\nu) < v(\sigma,\rho) + \epsilon. \tag{3.138}$$

This combined with $\ell(\nu) \ge \ell(0) = v(\sigma,\rho)$ gives $|\ell(\nu) - \ell(0)| < \epsilon$, thus establishing continuity at $\nu_0 = 0$. $\square$

Using Theorem 3.8.1, we establish the following asymptotic capacity result:

Theorem 3.8.10. *The capacity C of an AWGN channel with (σ,ρ)-power constraints and noise power ν satisfies the following:*

1. *When the noise power $\nu \to 0$, capacity C is given by*

$$C = v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu + \epsilon(\nu),$$

where $\epsilon(\cdot)$ is a function such that $\lim_{\nu\to 0}\epsilon(\nu) = 0$.

2. *When noise power $\nu \to \infty$, capacity C is given by*

$$C = \frac{1}{2}\left(\frac{\rho}{\nu}\right) - \frac{1}{4}\left(\frac{\rho}{\nu}\right)^2 + \frac{1}{6}\left(\frac{\rho}{\nu}\right)^3 + O\left(\left(\frac{\rho}{\nu}\right)^4\right).$$

Proof of Theorem 3.8.10. Note that all the logarithms used in this proof are taken to be at base e.

1. Using the lower bound in Theorem 3.3.2,

$$C \geq \frac{1}{2}\log\left(1 + \frac{e^{2v(\sigma,\rho)}}{2\pi e\nu}\right) \tag{3.139}$$

$$= v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu + \log\left(1 + \frac{2\pi e\nu}{e^{2v(\sigma,\rho)}}\right) \tag{3.140}$$

$$= v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu + O(\nu). \tag{3.141}$$

By continuity of ℓ at 0, we have that as $\nu \to 0$

$$\ell(\nu) = v(\sigma,\rho) + \epsilon(\nu) \tag{3.142}$$

for some $\epsilon(\cdot)$ satisfying $\lim_{\nu\to 0}\epsilon(\nu) = 0$. This gives the upper bound

$$C \leq v(\sigma,\rho) - \frac{1}{2}\log 2\pi e\nu + \epsilon(\nu). \tag{3.143}$$

Our claim follows from the inequalities (3.141) and (3.143). Unlike the case of $\sigma = 0$, we are unable to give any precise rate at which $\epsilon(\nu)$ goes to 0. Since we don't know what the intrinsic volumes of $\mathcal{S}_n(\sigma,\rho)$ are, we can only say that $-\Lambda^*(1-\theta)$ is continuous at $\theta = 0$, while not knowing how fast it approaches $v(\sigma,\rho)$ as $\theta \to 0$.

2. Note that C is bounded from below by the capacity of an AWGN channel with an amplitude constraint of $\sqrt{\rho}$. Using Theorem 3.7.6 we obtain for $\nu \to \infty$,

$$C \geq \frac{1}{2}\left(\frac{\rho}{\nu}\right) - \frac{1}{4}\left(\frac{\rho}{\nu}\right)^2 + \frac{1}{6}\left(\frac{\rho}{\nu}\right)^3 + O\left(\left(\frac{\rho}{\nu}\right)^4\right). \tag{3.144}$$

In addition, the upper bound from Theorem 3.3.2 states that

$$C \leq \frac{1}{2}\log\left(1 + \frac{\rho}{\nu}\right) = \frac{1}{2}\left(\frac{\rho}{\nu}\right) - \frac{1}{4}\left(\frac{\rho}{\nu}\right)^2 + \frac{1}{6}\left(\frac{\rho}{\nu}\right)^3 + O\left(\left(\frac{\rho}{\nu}\right)^4\right). \tag{3.145}$$

The claim now follows from equations (3.144) and (3.145).

□

3.9 Capacity results for general power constraints

In Sections 3.7 and 3.8 we studied the amplitude and (σ, ρ)-constraints respectively. In this section, we describe a general framework which can be used to analyze power-constrained Gaussian channels. We analyze two types of power constraints within this framework. The first type of constraint is what we call a "block constraint", which is essentially a vector generalization of the amplitude constraint. The second type of constraint is the "super-convolutive constraint", a natural constraint to encounter which includes the average power constraint as a special case.

3.9.1 Block constraints

Consider a channel with additive white Gaussian noise $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \nu I_d)$, where I_d is the $d \times d$ identity matrix. Let $K_d \subseteq \mathbb{R}^d$ be a compact convex set. The channel input denoted by $\mathbf{X} = (X_1, X_2, \cdots, X_d)$ is subject to the constraint

$$\mathbf{X} \in K_d \text{ almost surely.} \tag{3.146}$$

Let the capacity of this channel be C. For the amplitude A constraint, $d = 1$ and $K_1 = [-A, A]$. The quadrature Gaussian channel studied by Shamai & Bar-David [32] can also be studied in this setup. The power constraint therein can be described by choosing $d = 2$, and K_2 as the circle of radius $A > 0$. The main results of this section are Theorems 3.9.1, 3.9.2, and 3.9.3.

Theorem 3.9.1. *The capacity C satisfies the lower bound*

$$C \geq \frac{d}{2} \log \left(1 + \frac{e^{\frac{2}{d} \log Vol(K_d)}}{2\pi e \nu} \right). \tag{3.147}$$

Theorem 3.9.2. *Denote the intrinsic volumes of K_d by $\{\alpha_0, \cdots, \alpha_d\}$. Define $\Lambda(t)$ as*

$$\Lambda(t) = \log \left(\sum_{j=0}^{d} \alpha_j e^{jt} \right). \tag{3.148}$$

Let Λ^* be the convex conjugate of Λ. Then C satisfies the upper bound

$$C \leq \sup_{\theta} \left[-\Lambda^*(d\theta) + \frac{d(1-\theta)}{2} \log \frac{2\pi e \nu}{1-\theta} \right] - \frac{d}{2} \log 2\pi e \nu. \tag{3.149}$$

Theorem 3.9.3. *The asymptotic capacity as $\nu \to 0$ is given by*

$$C = \log Vol(K_d) - \frac{d}{2} \log 2\pi e \nu + O(\nu^{1/3}). \tag{3.150}$$

Proof of Theorem 3.9.1. Let $\mathcal{F}_d$ be the set of all distributions supported a.s. on K_d. Let $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ denote the channel output. We have

$$C = \sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} I(\mathbf{X}; \mathbf{Y}) \tag{3.151}$$

$$= \sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} h(\mathbf{Y}) - h(\mathbf{Z}) \tag{3.152}$$

$$= \sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} h(\mathbf{Y}) - \frac{d}{2} \log 2\pi e\nu \tag{3.153}$$

Using Shannon's entropy power inequality, we have

$$e^{\frac{2h(\mathbf{Y})}{d}} \geq e^{\frac{2h(\mathbf{X})}{d}} + e^{\frac{2h(\mathbf{Z})}{d}}. \tag{3.154}$$

Thus,

$$\begin{aligned} \sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} e^{\frac{2h(\mathbf{Y})}{d}} &\geq \sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} e^{\frac{2h(\mathbf{X})}{d}} + 2\pi e\nu \\ &= e^{\frac{2 \log \mathrm{Vol}(K_d)}{d}} + 2\pi e\nu. \end{aligned}$$

Taking logarithms on both sides, we have

$$\sup_{p_{\mathbf{X}}(\mathbf{X}) \in \mathcal{F}_d} h(\mathbf{Y}) \geq \frac{d}{2} \log \left(e^{\frac{2 \log \mathrm{Vol}(K_d)}{d}} + 2\pi e\nu \right), \tag{3.155}$$

which combined with inequality (3.153) concludes the proof. □

Proof of Theorem 3.9.2. Define $\ell(\nu)$ as

$$\ell(\nu) = \limsup_{n \to \infty} \frac{1}{n} \log \mathrm{Vol}(K_d^n \oplus B_{nd}(\sqrt{nd\nu}\,)), \tag{3.156}$$

where K_d^n is the n-product $K_d \times K_d \times \cdots \times K_d$. We can easily get the analogue of Theorem 3.6.1 and conclude that

$$C \leq \lim_{\epsilon \to 0} \ell(\nu + \epsilon) - \frac{d}{2} \log 2\pi e\nu. \tag{3.157}$$

Denote the intrinsic volumes of K_d^n by $\mu_{nd}(\cdot)$. Note that

$$\mu_{nd} = \underbrace{\mu_d \star \cdots \star \mu_d}_{n \text{ times}}. \tag{3.158}$$

We can use Steiner's formula to get

$$\mathrm{Vol}(K_d^n \oplus B_{nd}(\sqrt{nd\nu}\,)) = \sum_{j=0}^{nd} \mu_{nd}(nd - j)\epsilon_j \sqrt{nd\nu}^{\,j}. \tag{3.159}$$

Define the functions $a_n(\theta)$ and $b_n^\nu(\theta)$ for $\theta \in [0, d]$ as follows. The function $a_n(\theta)$ is obtained by linearly interpolating the values of $a_n(j/n)$, where the value of $a_n(j/n)$ is given by

$$a_n\left(\frac{j}{n}\right) = \frac{1}{n}\log \mu_{nd}(nd - j) \quad \text{for} \quad 0 \leq j \leq nd. \tag{3.160}$$

The function $b_n^\nu(\theta)$ is given by

$$b_n^\nu(\theta) = \frac{1}{n}\log \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2 + 1)}(nd\nu)^{n\theta/2} \quad \text{for} \quad \theta \in [0, d]. \tag{3.161}$$

Define $f_{nd}^\nu : [0, 1] \to \mathbb{R}$ as

$$f_n^\nu(\theta) := f^\nu(n, \theta) = a_n(\theta) + b_n^\nu(\theta). \tag{3.162}$$

With this notation, we can rewrite equation (3.159) as

$$\mathrm{Vol}(K_d^n \oplus B_{nd}(\sqrt{nd\nu}\,)) = \sum_{j=0}^{nd} e^{nf^\nu(nd, j/nd)}. \tag{3.163}$$

Using the same technique as in Lemma 3.8.4, we conclude that the functions a_n, b_n^ν, and f_n^ν are all concave. Define a sequence of measures supported on $[0, d]$ by

$$\mu_{nd/n}\left(\frac{j}{n}\right) := \mu_{nd}(j) \quad \text{for} \quad 0 \leq j \leq nd. \tag{3.164}$$

Using equation (3.158), we can directly apply Gärtner-Ellis theorem to conclude that the measures $\{\mu_{nd/n}\}$ converge in the large deviation-sense to $-\Lambda^*$. This is analogous to Lemma 3.8.5. The concavity of a_n along with the convergence of $\mu_{nd/n}$ in the large deviation sense to $-\Lambda^*$ can be used to show that $a_n(\theta)$ converges uniformly to $-\Lambda^*(d - \theta)$ exactly as in part 2 of Lemma 3.8.6. Using the same method as in part 1 of Lemma 3.8.6, we can also conclude that b_n^ν converges uniformly to $\frac{\theta}{2}\log\frac{2\pi ed\nu}{\theta}$. Thus, we have

$$f_n^\nu \text{ converges uniformly to } -\Lambda^*(d - \theta) + \frac{\theta}{2}\log\frac{2\pi ed\nu}{\theta}. \tag{3.165}$$

With this uniform convergence in hand, the analogues of Lemmas 3.8.7 and 3.8.8 readily follow and we conclude that

$$\ell(\nu) = \sup_{\theta \in [0,d]} -\Lambda^*(d - \theta) + \frac{\theta}{2}\log\frac{2\pi ed\nu}{\theta} \tag{3.166}$$

$$= \sup_{\theta \in [0,1]} -\Lambda^*(d\theta) + \frac{d(1-\theta)}{2}\log\frac{2\pi e\nu}{1-\theta}. \tag{3.167}$$

The continuity of ℓ can be established via the methods used in Theorem 3.8.1 and Lemma 3.8.9, to arrive at the upper bound

$$C \leq \sup_{\theta \in [0,1]} \left[-\Lambda^*(d\theta) + \frac{d(1-\theta)}{2} \log \frac{2\pi e \nu}{1-\theta} \right] - \frac{d}{2} \log 2\pi e \nu. \tag{3.168}$$

This concludes the proof of Theorem 3.9.2. $\square$

Proof of Theorem 3.9.3. Define

$$\theta^*(\nu) := \arg\sup_{\theta} \left[-\Lambda^*(d\theta) + \frac{d(1-\theta)}{2} \log \frac{2\pi e \nu}{1-\theta} \right]. \tag{3.169}$$

We shall sometimes refer to $\theta^*(\nu)$ simply as θ^* when the argument is understood. Note that $-\Lambda^*(d) = \log \mathrm{Vol}(K_d)$ and to show Theorem 3.9.3 it is enough to show that as $\nu \to 0$,

$$\limsup_{\nu \to 0} \left| \Lambda^*(d) + \left[-\Lambda^*(d\theta^*) + \frac{d(1-\theta^*)}{2} \log \frac{2\pi e \nu}{1-\theta^*} \right] \right| = O(\nu^{1/3}). \tag{3.170}$$

We rewrite this slightly as

$$\left| \left[\Lambda^*(d) - \Lambda^*(d\theta^*) + d(1-\theta^*) \log(1-\theta^*) \right] + \frac{d(1-\theta^*)}{2} \log \frac{2\pi e \nu}{(1-\theta^*)^3} \right|. \tag{3.171}$$

We break up the proof on Theorem 3.9.3 into two parts. The first part is contained in Lemma 3.9.4 below.

Lemma 3.9.4 (Proof in Appendix C.5.1)**.** *As $\nu \to 0$, the following equality holds:*

$$|\Lambda^*(d) - \Lambda^*(d\theta^*) + d(1-\theta^*) \log(1-\theta^*)| = O(\nu^{1/3}). \tag{3.172}$$

The second part claims the following:

Lemma 3.9.5 (Proof in Appendix C.5.2)**.** *As $\nu \to 0$, the following equality holds:*

$$\left| \frac{d(1-\theta^*)}{2} \log \frac{2\pi e \nu}{(1-\theta^*)^3} \right| = O(\nu^{1/3}). \tag{3.173}$$

Theorem 3.9.3 now follows from these two lemmas. $\square$

3.9.2 Super-convolutive constraints

Power constraints imposed on channel inputs can be described in a very general manner by considering a sequence of convex sets $\mathcal{K} = \{K_n\}_{n \geq 1}$ such that $K_n \subseteq \mathbb{R}^n$ for all n.

For all n, a sequence $(x_1, x_2, \cdots, x_n)$ is said to satisfy the power constraints imposed by $\{K_n\}$ if

$$(x_1, x_2, \cdots, x_n) \in K_n. \tag{3.174}$$

We shall refer to this as *power constrained by* $\{K_n\}$. We can describe the familiar constraints of average power and peak-power constraints with a suitable choice of $\{K_n\}$. For a peak power constraint of A, we choose K_n to be the n-dimensional cube $[-A, A]^n$ and for an average power constraint of P, we choose $K_n = B_n(\sqrt{nP})$ where $B_n(\sqrt{nP})$ is the Euclidean ball of radius $\sqrt{nP}$. In this section, we focus on families $\{K_n\}$ which are *super-convolutive*, defined as:

Definition. A sequence of sets $\{K_n\}$ is said to be super-convolutive if for all $m, n \geq 1$,

$$K_m \times K_n \subseteq K_{m+n}. \tag{3.175}$$

The power constraints imposed by such a sequence is called a *super-convolutive constraint*. A super-convolutive constraint is a natural kind of constraint to consider since it essentially states that if $(x_1, x_2, \cdots, x_n)$ and $(y_1, y_2, \cdots, y_m)$ are permissible codewords, then so is the concatenation $(x_1, \cdots, x_n, y_1, \cdots, y_m)$. The aforementioned examples of peak power and average power constraints are examples of super-convolutive constraints. Let the intrinsic volumes of K_n be denoted by $\mu_n(\cdot)$. The containment $K_m \times K_n \subseteq K_{m+n}$ implies

$$(\mu_m \star \mu_n)(j) \leq \mu_{m+n}(j) \quad \text{for all} \quad 0 \leq j \leq m+n. \tag{3.176}$$

Thus, the sequence of intrinsic volumes $\{\mu_n\}$ is a *super-convolutive sequence*. Such sequences have been studied in detail in Section 2.2.

We now make some assumptions on the family $\{K_n\}$, to ensure that the power constraints imposed by $\{K_n\}$ are reasonable. The assumptions are as follows:

$$(\mathbf{P}): \; K_1 \neq \phi \tag{3.177}$$

$$(\mathbf{Q}): \; \lim_n \frac{1}{n} \log \mathrm{Vol}(K_n) = \alpha < \infty \tag{3.178}$$

$$(\mathbf{R}): \; \lim_n \sum_j \mu_n(j) = \gamma < \infty \tag{3.179}$$

Assumption $(\mathbf{P})$ ensures that K_n is an n-dimensional set, since $K_1^n \subseteq K_n$. Thus, for all n, the intrinsic volumes $\mu_n(0)$ and $\mu_n(n)$ satisfy

$$\mu_n(0) = 1, \quad \text{and} \quad \mu_n(n) > 0. \tag{3.180}$$

From inequality (3.176) we have

$$\mu_n(n)\mu_m(m) \geq \mu_{m+n}(m+n), \tag{3.181}$$

which implies

$$\mathrm{Vol}(K_n)\mathrm{Vol}(K_m) \leq \mathrm{Vol}(K_{m+n}). \tag{3.182}$$

Using Fekete's lemma, we obtain that the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(K_n) \text{ exists, and is possibly } +\infty. \tag{3.183}$$

Assumption (**Q**) states that this limit must be finite. Assumption (**R**) appears technical, however it can be easily seen to be satisfied in most cases of interest, as follows.

Define γ_n as

$$\gamma_n = \log \sum_j \mu_n(j). \tag{3.184}$$

Upon summing both sides of inequality (3.176) over all j, we get

$$\gamma_{m+n} \geq \gamma_n + \gamma_m. \tag{3.185}$$

Thus, the limit

$$\lim_{n\to\infty} \frac{\gamma_n}{n} \tag{3.186}$$

exists, and is possibly $+\infty$. To rule out this limit being $+\infty$, it is enough to show a finite upper bound on this limit. One way to establish such an upper bound is to search for a large enough $R > 0$ such that for all n,

$$K_n \subseteq B_n(\sqrt{nR}). \tag{3.187}$$

This simply means that for some finite R, there is an average power constraint of R on the transmitted codewords. Let us denote the intrinsic volumes of $B_n(\sqrt{nR})$ by $\{\hat{\mu}_n(\cdot)\}$. The containment (3.187) implies $\mu_n \leq \hat{\mu}_n$. Thus,

$$\log \sum_j \mu_n(j) \leq \log \sum_j \hat{\mu}_n(j). \tag{3.188}$$

We divide both sides by n and take the limit as $n \to \infty$. Note that the limit of the right hand side can by explicitly evaluated to equal a finite $\hat{\gamma}$, and we can conclude

$$\gamma \leq \hat{\gamma} < \infty. \tag{3.189}$$

We can easily check that if the assumptions $(\mathbf{P}), (\mathbf{Q})$, and $(\mathbf{R})$ are satisfied, the sequence of intrinsic volumes $\{\mu_n(\cdot)\}$ satisfies the assumptions $(\mathbf{A}), (\mathbf{B}), (\mathbf{C})$, and $(\mathbf{D})$ from Section 2.2. Thus, all the convergence results proved for super-convolutive sequences in Section 2.2 apply directly for the sequence of intrinsic volumes.

Henceforth, $\{K_n\}$ is understood to be a super-convolutive sequence satisfying the assumptions $(\mathbf{P}), (\mathbf{Q})$ and $(\mathbf{R})$. Consider a scalar additive Gaussian channel with noise power ν, and input power constrained by $\{K_n\}$. Let the capacity of this channel be C. As in Section 2.2, define

$$G_n(t) = \log \sum_j \mu_n(j) e^{jt}, \quad g_n(t) = \frac{G_n(t)}{n}, \quad \text{and} \quad \Lambda(t) = \lim_n g_n(t). \tag{3.190}$$

Let Λ^* be the convex conjugate of Λ. The main results of this section are as follows:

Theorem 3.9.6. *The capacity C is bounded from above as*

$$C \leq \sup_\theta \left[-\Lambda^*(1-\theta) + \frac{\theta}{2} \log \frac{2\pi e \nu}{\theta} \right] - \frac{1}{2} \log 2\pi e \nu. \tag{3.191}$$

Theorem 3.9.7. *The capacity C is bounded from below as*

$$C \geq \frac{1}{2} \log \left(1 + \frac{e^{-2\Lambda^*(1)}}{2\pi e \nu} \right). \tag{3.192}$$

Theorem 3.9.8. *The asymptotic capacity C as $\nu \to 0$ is given by*

$$C = -\Lambda^*(1) - \frac{1}{2} \log 2\pi e \nu + \epsilon(\nu), \tag{3.193}$$

where the error term $\epsilon(\nu) \to 0$ as $\nu \to 0$.

At first glance, these results seem similar to those obtained in Section 3.8 and sub-section 3.9.1. However, there is an important difference. In the results for our previous examples, in place of the $-\Lambda^*(1)$ term we had the exponential growth-rate of the volume of $\{K_n\}$; $\log 2A$ for the peak power constraint, $v(\sigma, \rho)$ for the (σ, ρ)-constraint, $\log \mathrm{Vol}(K_d)$ for the block constraint, and $\frac{1}{2} \log 2\pi e P$ for the average power constraint. One reason for this is that in all the cases previously encountered, $-\Lambda^*(1)$ equalled the exponential growth rate of volume. It is natural to wonder if the equality

$$-\Lambda^*(1) = \lim_n \frac{1}{n} \log \mathrm{Vol}(K_n) \tag{3.194}$$

holds for the case of super-convolutive constraints. In case it holds, the results of this section do indeed exactly parallel those of Sections 3.8 and 3.9.1. We as yet do not have a example of a super-convolutive family $\{K_n\}$ in which the equality (3.194) doesn't hold. However, we believe that it is likely that such an example exists. This is supported by the example in Section 2.2 where we construct a super-convolutive sequence $\{\mu_n\}$ for which

$$-\Lambda^*(1) \neq \lim_{n \to \infty} \frac{1}{n} \log \mu_n(n). \tag{3.195}$$

Note that the constructed sequence does not correspond to the intrinsic volumes of any convex sets $\{K_n\}$, and therefore does not provide a counterexample to the equality in (3.194).

Our results imply that in case equality doesn't hold, capacity depends not on the exponential growth rate of volume as intuition might suggest, but on the value of $-\Lambda^*(1)$.

Proof of Theorem 3.9.6. Define $\ell(\nu)$ as

$$\ell(\nu) = \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(K_n \oplus B_n(\sqrt{n\nu}\,)), \tag{3.196}$$

We can easily get the analogue of Theorem 3.6.1 and conclude that

$$C \leq \lim_{\epsilon\to 0} \ell(\nu+\epsilon) - \frac{1}{2}\log 2\pi e\nu. \tag{3.197}$$

By Steiner's formula, we have

$$\mathrm{Vol}(K_n \oplus B_n(\sqrt{n\nu}\,)) = \sum_{j=0}^{n} \mu_n(n-j)\epsilon_j \sqrt{n\nu}^{\,j}. \tag{3.198}$$

Define the functions $a_n(\theta)$ and $b_n^\nu(\theta)$ for $\theta \in [0,1]$ as follows. The function $a_n(\theta)$ is obtained by linearly interpolating the values of $a_n(j/n)$, where the value of $a_n(j/n)$ is given by:

$$a_n\left(\frac{j}{n}\right) = \frac{1}{n}\log \mu_n(n-j) \quad \text{for} \quad 0 \leq j \leq n. \tag{3.199}$$

The function $b_n^\nu(\theta)$ is given by

$$b_n^\nu(\theta) = \frac{1}{n}\log \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)}(n\nu)^{n\theta/2} \quad \text{for} \quad \theta \in [0,1]. \tag{3.200}$$

Define $f_n^\nu : [0,1] \to \mathbb{R}$ as

$$f_n^\nu(\theta) := f^\nu(n,\theta) = a_n(\theta) + b_n^\nu(\theta). \tag{3.201}$$

With this notation, we can rewrite equation (3.198) as

$$\mathrm{Vol}(K_n \oplus B_n(\sqrt{n\nu})) = \sum_{j=0}^{n} e^{nf^\nu(n,j/n)}. \tag{3.202}$$

Using the same technique as in Lemma 3.8.4, we conclude that the functions a_n, b_n^ν, and f_n^ν are all concave. Define a sequence of measures supported on $[0,1]$ by

$$\mu_{n/n}\left(\frac{j}{n}\right) := \mu_n(j) \quad \text{for} \quad 0 \leq j \leq n. \tag{3.203}$$

Theorems 2.2.2 and 2.2.4 imply that the measures $\{\mu_{n/n}\}$ converge in the large deviation-sense to $-\Lambda^*$. This is analogous to Lemma 3.8.5. The concavity of a_n along with the convergence of $\mu_{n/n}$ in the large deviation sense to $-\Lambda^*$ can be used to show that $a_n(\theta)$ converges pointwise to $-\Lambda^*(1-\theta)$ on the open interval $(0,1)$ exactly as in part 2 of Lemma 3.8.6. However, since convergence at the endpoints is not known, we can no longer obtain any results about uniform convergence of a_n as in Lemma 3.8.6. Therefore, our proof deviates slightly from that of Theorems 3.8.1 and 3.9.2.

Denote

$$\mathcal{A}^\nu(\theta) := -\Lambda^*(1-\theta) + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}, \tag{3.204}$$

and

$$A(\nu) := \sup_{\theta\in[0,1]} \mathcal{A}^\nu(\theta). \tag{3.205}$$

The continuity of A can be readily established via the methods used in Theorem 3.8.1 and Lemma 3.8.9. We prove the following lemma:

Lemma 3.9.9 (Proof in Appendix C.5.3)**.** *For any $\eta > 0$, there exists an N such that for all $n > N$,*

$$f_n^\nu(\theta) < \mathcal{A}^\nu(\theta) + \eta, \tag{3.206}$$

for all $\theta \in [0,1]$.

Let $\hat{\theta}_n = \arg\max_{j/n} f_n^\nu(j/n)$. Using the same analysis as in Lemma 3.8.8, we obtain

$$\ell(\nu) = \limsup_n f^\nu(n, \hat{\theta}_n). \tag{3.207}$$

Let $\eta > 0$ be given. Choose N according to Lemma 3.9.9 such that

$$f_n^\nu < \mathcal{A}^\nu + \eta/2 \quad \text{for} \quad n > N. \tag{3.208}$$

Using the continuity of A, choose $\epsilon > 0$ small enough such that $A(\nu+\epsilon) < A(\nu) + \eta/2$. We have the sequence of inequalities

$$\begin{align}
\ell(\nu+\epsilon) &= \limsup_n f_n^{\nu+\epsilon}(\hat{\theta}_n) \tag{3.209}\\
&< \limsup_n \mathcal{A}^{\nu+\epsilon}(\hat{\theta}_n) + \eta/2 \tag{3.210}\\
&\le \limsup_n \left[\sup_\theta \mathcal{A}^{\nu+\epsilon}(\theta)\right] + \eta/2 \tag{3.211}\\
&= A(\nu+\epsilon) + \eta/2 \tag{3.212}\\
&< A(\nu) + \eta. \tag{3.213}
\end{align}$$

Thus,

$$\lim_{\epsilon\to 0} \ell(\nu+\epsilon) < A(\nu) + \eta. \tag{3.214}$$

This gives us that for any $\eta > 0$, the capacity C is bounded from above by

$$C \leq A(\nu) + \eta - \frac{1}{2}\log 2\pi e\nu. \tag{3.215}$$

Letting $\eta \to 0$, we conclude the proof. □

Proof of Theorem 3.9.7. Establishing a lower bound on capacity requires an achievable scheme. If we use the same scheme as in Theorem 3.3.2, we get the lower bound

$$C \geq \frac{1}{2}\log\left(1 + \frac{e^{2\alpha}}{2\pi e\nu}\right), \tag{3.216}$$

where $\alpha = \lim_n \frac{1}{n}\log \mathrm{Vol}(K_n)$. However, by Lemma 2.2.3, we have the inequality $-\Lambda^*(1) \geq \alpha$. Thus, this bound is weaker than the bound in Theorem 3.9.7, and we need to devise a slightly different achievable scheme.

Fix a $1 > \theta_0 > 0$. Suppose we pick a $\lfloor n\theta_0 \rfloor$ dimensional subspace of $\mathbb{R}^n$ uniformly at random, and take the projection of K_n on this subspace. The results from Klain & Rota [20] imply that the mean volume of the projection is related to the intrinsic volumes of K_n as follows:

$$\text{Mean volume of projection} = \mu_{\lfloor n\theta \rfloor}(K_n)\frac{\epsilon_{\lfloor n\theta \rfloor}\epsilon_{n-\lfloor n\theta \rfloor}}{\epsilon_n}\frac{1}{\binom{n}{\lfloor n\theta \rfloor}} \tag{3.217}$$

where ϵ_j is the volume of the j-dimensional unit ball. This means that for every K_n, there exists an $\lfloor n\theta \rfloor$ dimensional subspace V such that

$$\mathrm{Vol}_{\lfloor n\theta \rfloor}(K_n \perp V) \geq \mu_{\lfloor n\theta_0 \rfloor}(K_n)\frac{\epsilon_{\lfloor n\theta_0 \rfloor}\epsilon_{n-\lfloor n\theta_0 \rfloor}}{\epsilon_n}\frac{1}{\binom{n}{\lfloor n\theta_0 \rfloor}}. \tag{3.218}$$

Let W be $V^\perp$. Every point $x^n \in K_n$ can be expressed in terms of its orthogonal projections on V and W as, say $x_V^n + x_W^n$. Choose a distribution of X^n supported on K_n such that X_V^n is uniformly distributed on $K_n \perp V_n$. We have the lower bound on capacity,

$$\begin{aligned} nC &\geq I(X^n; Y^n) && (3.219)\\ &\geq I(X_V^n; Y_V^n) && (3.220)\\ &= h(Y_V^n) - \frac{\lfloor n\theta_0 \rfloor}{2}\log 2\pi e\nu. && (3.221) \end{aligned}$$

This gives

$$C \geq \frac{h(Y_V^n)}{n} - \frac{\lfloor n\theta_0 \rfloor}{2n}\log 2\pi e\nu, \tag{3.222}$$

which upon taking a lim sup implies

$$C \geq \limsup_n \frac{h(Y_V^n)}{n} - \frac{\lfloor n\theta_0 \rfloor}{2n} \log 2\pi e \nu \tag{3.223}$$

$$= \limsup_n \frac{h(Y_V^n)}{n} - \frac{\theta_0}{2} \log 2\pi e \nu. \tag{3.224}$$

An application of the entropy power inequality gives

$$e^{\frac{2h(Y_V^n)}{\lfloor n\theta_0 \rfloor}} \geq e^{\frac{2h(X_V^n)}{\lfloor n\theta_0 \rfloor}} + e^{\frac{2h(Z_V^n)}{\lfloor n\theta_0 \rfloor}}, \tag{3.225}$$

which implies

$$\frac{h(Y_V^n)}{n} \geq \frac{\lfloor n\theta_0 \rfloor}{2n} \log \left(e^{\frac{2 \log \mathrm{Vol}(K_n \perp V_n)}{\lfloor n\theta_0 \rfloor}} + 2\pi e \nu \right) \tag{3.226}$$

We now take the limit as $n \to \infty$ and evaluate the right hand side. This essentially boils down to computing the limit

$$\lim_n \frac{2 \log \mathrm{Vol}(K_n \perp V)}{\lfloor n\theta_0 \rfloor}.$$

By the choice of V, we have from inequality (3.218)

$$\frac{1}{n} \log \mathrm{Vol}(K_n \perp V) \geq \frac{1}{n} \log \left(\mu_{\lfloor n\theta_0 \rfloor}(K_n) \frac{\epsilon_{\lfloor n\theta_0 \rfloor} \epsilon_{n - \lfloor n\theta_0 \rfloor}}{\epsilon_n} \frac{1}{\binom{n}{\lfloor n\theta_0 \rfloor}} \right) \tag{3.227}$$

$$= \frac{1}{n} \log \mu_{\lfloor n\theta_0 \rfloor}(K_n) + \frac{1}{n} \log \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{\lfloor n\theta_0 \rfloor}{2}+1)\Gamma(\frac{n-\lfloor n\theta_0 \rfloor}{2}+1)} - \frac{1}{n} \log \binom{n}{\lfloor n\theta_0 \rfloor} \tag{3.228}$$

Taking the limit as $n \to \infty$ and using the pointwise convergence of $a_n(\theta)$ defined in equation (3.199), we arrive at

$$\lim_n \frac{1}{n} \log \mathrm{Vol}(K_n \perp V) \geq -\Lambda^*(\theta_0) + \frac{H(\theta_0)}{2} - H(\theta_0) \tag{3.229}$$

$$= -\Lambda^*(\theta_0) - \frac{H(\theta_0)}{2}. \tag{3.230}$$

Substituting in inequality (3.226),

$$\limsup_n \frac{h(Y_V^n)}{n} \geq \frac{\theta_0}{2} \log \left(e^{-\frac{2}{\theta_0}\Lambda^*(\theta_0) - \frac{H(\theta_0)}{\theta_0}} + 2\pi e \nu \right). \tag{3.231}$$

This gives the lower bound on capacity,

$$C \geq \frac{\theta_0}{2} \log \left(1 + \frac{e^{-\frac{2}{\theta_0}\Lambda^*(\theta_0) - \frac{H(\theta_0)}{\theta_0}}}{2\pi e \nu} \right). \tag{3.232}$$

Since this lower bound holds for any choice of $\theta_0 < 1$, we can take the limit as $\theta_0 \to 1$ and conclude

$$C \geq \frac{1}{2} \log \left(1 + \frac{e^{-2\Lambda^*(1)}}{2\pi e \nu}\right). \tag{3.233}$$

□

Proof of Theorem 3.9.8. Theorem 3.9.6 establishes the upper bound

$$C \leq A(\nu) - \frac{1}{2} \log 2\pi e \nu. \tag{3.234}$$

Note that $A(0) = -\Lambda^*(1)$ and A is continuous at $\nu = 0$. Thus, as $\nu \to 0$ we can write the upper bound as

$$C \leq -\Lambda^*(1) - \frac{1}{2} \log 2\pi e \nu + \epsilon_1(\nu), \tag{3.235}$$

where $\epsilon_1(\nu) \to 0$ as $\nu \to 0$.

Theorem 3.9.7 establishes the lower bound

$$C \geq \frac{1}{2} \log \left(1 + \frac{e^{-2\Lambda^*(1)}}{2\pi e \nu}\right) \tag{3.236}$$

$$= -\Lambda^*(1) - \frac{1}{2} \log 2\pi e \nu + \frac{1}{2} \log \left(1 + \frac{2\pi e \nu}{e^{-2\Lambda^*(1)}}\right) \tag{3.237}$$

$$= -\Lambda^*(1) - \frac{1}{2} \log 2\pi e \nu + \epsilon_2(\nu) \tag{3.238}$$

where $\epsilon_2(\nu) \to 0$ as $\nu \to 0$.

Equations (3.235) and (3.238) lead to the conclusion that as $\nu \to 0$, the capacity C is given by

$$C = -\Lambda^*(1) - \frac{1}{2} \log 2\pi e \nu + \epsilon(\nu), \tag{3.239}$$

for an error term $\epsilon(\nu)$ which tends to 0 as ν tends to 0. □

3.10 Conclusion

In this chapter, we studied in detail an AWGN channel with a power constraint motivated by energy harvesting communication systems, called the (σ, ρ)-power constraint. Such a power constraint induces an infinite memory in the channel. In general, finding capacity expressions for channels with memory is hard, even if we allow for n-letter capacity expressions. However, in this particular case, we are able to exploit the following geometric properties of $\{\mathcal{S}_n(\sigma, \rho)\}$:

A : $\mathcal{S}_{m+n}(\sigma, \rho) \subseteq \mathcal{S}_n(\sigma, \rho) \times \mathcal{S}_m(\sigma, \rho)$,

B : $[\mathcal{S}_m(\sigma, \rho) \times \mathbf{0}_k] \times [\mathcal{S}_n(\sigma, \rho) \times \mathbf{0}_k] \subseteq \mathcal{S}_{m+n+2k}(\sigma, \rho)$, when $k = \lceil \frac{\sigma}{\rho} \rceil$.

Property (**A**) allowed us to upper-bound channel capacity, and property (**B**) allowed us to lower-bound the same. In Section 3.2, we used these two properties to establish an n-letter capacity expression.

The main contribution of Section 3.3 was the EPI based lower bound. To arrive at this lower bound, we used the n-letter capacity expression from Section 3.2, and the following property:

C : The limit $\lim_n \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_n(\sigma, \rho))$ exists, and is finite.

For most reasonable power constraints, an exponential volume growth rate as defined in property **C** can be shown to exist. The case of (σ, ρ)-constraints was especially interesting, because it was fairly easy to evaluate $v(\sigma, \rho)$ using the numerical method in Section 3.5. We attribute this ease to the existence of a state σ_n, which is a single parameter that encapsulates all the relevant information about the history of the sequence. We used the computed value of $v(\sigma, \rho)$ to plot the EPI based lower bound. Our results show that energy harvesting communication systems have significant capacity gains even for a small battery. We then established an upper bound on capacity using the exponential growth rate of volume of the Minkowski sum of $\mathcal{S}_n(\sigma, \rho)$ and a ball of radius $\sqrt{n\nu}$. For the special case of $\sigma = 0$, which is the peak power constrained AWGN channel, we explicitly evaluated this upper bound. This enabled us to derive new asymptotic capacity results for such a channel. We also established a new upper bound on the entropy $h(X + Z)$, when X is amplitude-constrained, and Z is variance-constrained. The analysis for the case of $\sigma > 0$ was more involved because the intrinsic volumes of $\mathcal{S}_n(\sigma, \rho)$ are not known in a closed form. Using a new notion of sub-convolutive sequences, we showed that the logarithms of the intrinsic volumes of $\{\mathcal{S}_n(\sigma, \rho)\}$ when appropriately normalized, converge to a limit function. We then established an asymptotic capacity result in terms of this limit function. Our analysis crucially depended on both, property (**A**) and property (**B**).

In Section 3.9 we described a general framework to analyze power constrained AWGN channels, and analyzed the two special cases of block power constraints and super-convolutive power constraints. The block power constraint is essentially a vector version of the peak power constraint. We extended the techniques from Section 3.7 and established a capacity upper bound, and asymptotic capacity results for an AWGN channel with such constraints.

Note that for the (σ, ρ)-constraints, we had to rely on both properties (**A**) and (**B**) to prove the capacity upper bound. Many constraints, such as the average power constraint, do not satisfy property (**A**). We considered an AWGN channel with a super-convolutive constraint to study precisely such a scenario; when (**B**) alone is satisfied. We showed that not having property (**A**) is not a big handicap. We proved upper and lower bounds on capacity, as well as asymptotic capacity results for such a constraint. However, these bounds are not in terms of the exponential growth of volume, but in terms of $-\Lambda^*(1)$. For all the examples we considered, the term $-\Lambda^*(1)$ equalled the exponential growth rate of volume. As yet, we do not have an example of a superconvolutive power constraint where this equality doesn't hold, but we conjecture that it is possible to find such an example.

Chapter 4

Geometry of typical sets

Among the relations between information theory and geometry, the relation between differential entropy and volume is the most popular. Given a real-valued random variable X with density p_X and differential entropy $h(X)$, one way to define its ϵ-typical set, $\hat{\mathcal{T}}_n^\epsilon$ in dimension n, is

$$\{x^n \in \mathbb{R}^n \,|\, e^{-n(h(X)+\epsilon)} \leq p_{X^n}(x^n) \leq e^{-n(h(X)-\epsilon)}\}, \tag{4.1}$$

where $p_{X^n}(x^n) = \prod p_X(x_i)$. A well-known fact [7] is that for all large enough n, the volume, $|\hat{\mathcal{T}}_n^\epsilon|$, satisfies

$$(1-\epsilon)e^{n(h(X)-\epsilon)} \leq |\hat{\mathcal{T}}_n^\epsilon| \leq e^{n(h(X)+\epsilon)}. \tag{4.2}$$

Thus, the exponential growth rate of the volume $|\hat{\mathcal{T}}_n^\epsilon|$ is determined by the differential entropy $h(X)$. This connection between differential entropy and volume extends to inequalities. For instance, the Brunn-Minkowski inequality [14] states that any two compact convex sets $A, B \subseteq \mathbb{R}^n$ satisfy

$$|A|^{1/n} + |B|^{1/n} \leq |A \oplus B|^{1/n},$$

where $A \oplus B$ is the Minkowski sum [30] of A and B. This is strikingly similar to the entropy power inequality (EPI) [33], which states that any two independent $\mathbb{R}^n$-valued random variables $\mathbf{X}$ and $\mathbf{Y}$ satisfy

$$e^{2h(\mathbf{X})/n} + e^{2h(\mathbf{Y})/n} \leq e^{2h(\mathbf{X}+\mathbf{Y})/n}.$$

This connection has been explored in detail in [14] and [5].

Consider the sequence of typical sets $\{\hat{\mathcal{T}}_n^\epsilon\}_{n\geq 1}$ defined by the inequalities in (4.1). The bounds in (4.2) describe differential entropy in terms of a specific geometric function (the volume) of $\{\hat{\mathcal{T}}_n^\epsilon\}_{n\geq 1}$. It is also possible to consider other functions, apart from volume, for instance intrinsic volumes. It is natural to ask what the $\mathcal{G}$-function of typical sets is; i.e. how the intrinsic volumes of a sequence of typical sets such as $\{\hat{\mathcal{T}}_n^\epsilon\}$

scales with n. This question does not always make sense, since intrinsic volumes are not defined for arbitrary sets. We therefore consider the following setting: Let X be a real-valued random variable with a log-concave density $p_X(X) := e^{-\Phi(x)}$, for a convex function $\Phi : \mathbb{R} \to \mathbb{R} \cup +\infty$. For each $n \geq 1$ and $\epsilon > 0$, define the one-sided ϵ-typical set as follows:

$$\mathcal{T}_n^\epsilon = \mathrm{cl}\left(\{x^n \in \mathbb{R}^n \mid p_{X^n}(x^n) \geq e^{-n(h(X)+\epsilon)}\}\right) \tag{4.3}$$

$$= \mathrm{cl}\left(\{x^n \in \mathbb{R}^n \mid \sum_{i=1}^n \Phi(x_i) \leq n(h(X)+\epsilon)\}\right), \tag{4.4}$$

where cl stands for closure of a set. The definition of $\mathcal{T}_n^\epsilon$ and the convexity of $\sum_{i=1}^n \Phi(x_i)$ immediately imply that $\mathcal{T}_n^\epsilon$ is a compact convex set, hence belongs in $\mathcal{K}_n$. Let us denote the intrinsic volumes of $\mathcal{T}_n^\epsilon$ by $\{\mu_n^\epsilon(0), \dots, \mu_n^\epsilon(n)\}$. As noted earlier, the n^{th} intrinsic volume is simply the volume, and its exponential growth rate is determined by the differential entropy since (4.2) continues to hold for $\mathcal{T}_n^\epsilon$. This is stated as

$$h(X) = \lim_{\epsilon \to 0_+} \lim_{n \to \infty} \frac{1}{n} \log \mu_n^\epsilon(n). \tag{4.5}$$

We look at the limit

$$h_\theta(X) := \lim_{\epsilon \to 0_+} \lim_{n \to \infty} \frac{1}{n} \log \mu_n^\epsilon(\lfloor n\theta \rfloor), \tag{4.6}$$

for $\theta \in [0,1]$. Note that for $\theta = 1$, we have $h_1(X) = h(X)$ from equations (4.5) and (4.6). The value of $h_0(X)$ can be seen to equal 0, since the 0^{th} intrinsic volume (i.e. the Euler characteristic) is always 1 for non-empty compact convex sets [20]. For $\theta \in (0,1)$, the existence of the limit in equation (4.6) is not a priori obvious. For each value of $\theta \in [0,1]$, the quantity $h_\theta(X)$ provides the exponential growth rate of the $\lfloor n\theta \rfloor^{\text{th}}$ intrinsic volume of typical sets, and may be viewed as a generalization of differential entropy for log-concave distributions. We look at two examples where the limit $h_\theta(X)$ can be evaluated in a closed form:

Example 3. Let $X \sim \mathcal{N}(0,\nu)$. The one-sided ϵ-typical set in this case is simply the n-dimensional ball of radius $n\nu(1+2\epsilon)$, denoted by $B_n(\sqrt{n\nu(1+2\epsilon)})$. The intrinsic volumes such a ball admit a closed form expression [20], and the j^{th} intrinsic volume is given by

$$V_j\left(B_n(\sqrt{n\nu(1+2\epsilon)})\right) = \binom{n}{j} \frac{\omega_j}{\omega_{n-j}} \left(n\nu(1+2\epsilon)\right)^{j/2} \tag{4.7}$$

where ω_i is the volume of the i-dimensional unit ball. Substituting $j = \lfloor n\theta \rfloor$, and taking the desired limits yields

$$h_\theta(X) = H(\theta) + \frac{\theta}{2} \log 2\pi e\nu + \frac{1-\theta}{2} \log(1-\theta), \tag{4.8}$$

where $H(\theta) = -\theta \log \theta - (1-\theta)\log(1-\theta)$ is the binary entropy function.

Example 4. Let X be a random variable distributed uniformly in the interval $[0, A]$. For all $\epsilon > 0$, the one-sided ϵ-typical set for X is the n-dimensional cube $[0, A]^n$. The j^{th} intrinsic volume of this cube [20] is given by

$$V_j\left([0, A]^n\right) = \binom{n}{j} A^j. \tag{4.9}$$

Substituting $j = \lfloor n\theta \rfloor$ and taking the desired limits gives

$$h_\theta(X) = H(\theta) + \theta \log A. \tag{4.10}$$

For an arbitrary log-concave distribution, such an explicit calculation is not possible as the intrinsic volumes of its typical sets are not available in closed form. We will show that for all log-concave distributions, the limit $h_\theta(X)$ exists for each value of $\theta \in [0, 1]$ and $h_\theta(X)$ viewed as a function of θ is continuous on $[0, 1]$.

In Section 4.1, we show that the sequence of intrinsic volumes of $\{\mathcal{T}_n^\epsilon\}$ is superconvolutive and apply results from Section 2.2 to these sequences. In Section 4.2, we take the limit of $(\Lambda^\epsilon)^*$ as $\epsilon \to 0_+$ and show that the limit function Λ^* is a continuous, concave function, which equals h_θ.

4.1 Large deviations type convergence of intrinsic volumes

Let X be a real-valued random variable with density $p_X(X)$. We assume $p_X(X)$ is log-concave, hence is given by $p_X(x) = e^{-\Phi(x)}$ for a convex function $\Phi : \mathbb{R} \to \mathbb{R} \cup +\infty$. Note that $\int_{\mathbb{R}} e^{-\Phi(x)} dx = 1$, so $\Phi(x) \to +\infty$ as $x \to \pm\infty$.

Lemma 4.1.1. *The sequence of sets* $\{\mathcal{T}_n^\epsilon\}_{n \geq 1}$ *satisfies*

$$\mathcal{T}_m^\epsilon \times \mathcal{T}_n^\epsilon \subseteq \mathcal{T}_{m+n}^\epsilon, \quad \textit{for all} \;\; m, n \geq 1. \tag{4.11}$$

Proof. Let $x^m \in \text{int}(\mathcal{T}_m^\epsilon)$ and $y^n \in \text{int}(\mathcal{T}_n^\epsilon)$, where int stands for interior of a set. We have

$$\sum_{i=1}^{m} \Phi(x_i) \leq m(h(X) + \epsilon) \quad \text{and} \quad \sum_{i=1}^{n} \Phi(y_i) \leq n(h(X) + \epsilon).$$

Adding the above inequalities, $z^{m+n} = (x^m, y^n)$ satisfies

$$\sum_{i=1}^{m+n} \Phi(z_i) \leq (m + n)(h(X) + \epsilon),$$

which implies that $z^{m+n} \in \text{int}(\mathcal{T}_{m+n}^\epsilon)$. The result for boundary points follows by considering appropriate limiting sequences. □

As described in Chapter 1, the sequence of intrinsic volumes $\{\mu_n^\epsilon\}_{n\geq 1}$ forms a super-convolutive sequence. The convergence properties of these sequences have been studied in detail in Section 2.2. Before we state our main theorem for this section, we introduce some notation. Define

$$G_n^\epsilon(t) = \log \sum_{j=0}^{n} \mu_n^\epsilon(j) e^{jt}, \quad \text{and} \quad g_n^\epsilon(t) = \frac{G_n^\epsilon(t)}{n}. \tag{4.12}$$

The super-convolutivity of $\{\mu_n^\epsilon\}$ implies that

$$G_m^\epsilon(t) + G_n^\epsilon(t) \leq G_{m+n}^\epsilon(t), \quad \forall m, n \geq 1 \quad \text{and} \quad \forall t. \tag{4.13}$$

Thus, for each t the sequence $\{G_n^\epsilon(t)\}$ is super-additive, and by Fekete's lemma [35], the limit $\lim_n g_n^\epsilon(t)$ exists, though it is possibly $+\infty$. Define

$$\Lambda^\epsilon(t) := \lim_{n\to\infty} g_n^\epsilon(t), \tag{4.14}$$

and let $(\Lambda^\epsilon)^*$ be the convex conjugate [3] of Λ^ϵ.

We first check that the super-convolutive sequence of intrinsic volumes $\{\mu_n^\epsilon(\cdot)\}_{n\geq 1}$ satisfies properties $(\mathbf{A}), (\mathbf{B}), (\mathbf{C})$ and $(\mathbf{D})$ described in Section 2.2, so that we can use the convergence results contained therein.

Lemma 4.1.2 (Proof in Appendix D.1.1)**.** *For all $n \geq 1$, we have $\mu_n^\epsilon(0) > 0$ and $\mu_n^\epsilon(n) > 0$. Let $\alpha := \lim_n \frac{\log \mu_n^\epsilon(n)}{n}$, $\beta := \lim_n \frac{\log \mu_n^\epsilon(0)}{n}$, and $\gamma := \lim_n g_n^\epsilon(0)$. Then $\alpha, \beta, \gamma < \infty$.*

Applying Theorem 2.2.2 and Theorem 2.2.4 directly, we arrive at

Theorem 4.1.3. *Define a sequence of measures $\{\mu_{n/n}^\epsilon\}_{n\geq 1}$ supported on $[0,1]$ by*

$$\mu_{n/n}^\epsilon\left(\frac{j}{n}\right) := \mu_n^\epsilon(j), \quad \textit{for} \quad 0 \leq j \leq n.$$

Let $I \subseteq \mathbb{R}$ be a closed set and $F \subseteq \mathbb{R}$ be an open set. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_{n/n}^\epsilon(I) \leq -\inf_{x\in I} (\Lambda^\epsilon)^*(x), \qquad \textit{and} \tag{4.15}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}^\epsilon(F) \geq -\inf_{x\in F} (\Lambda^\epsilon)^*(x). \tag{4.16}$$

4.2 The limit function $-\Lambda^*$

In Section 4.1, we showed that the function $(\Lambda^\epsilon)^*$ plays the role of a large deviations rate function for the sequence of intrinsic volumes of $\{\mathcal{T}_n^\epsilon\}$. We now take the limit of $(\Lambda^\epsilon)^*$ as $\epsilon \to 0_+$, to obtain a limit function which is independent of ϵ and depends only on the starting density p_X. We show the following theorem:

Theorem 4.2.1. *Define the function* $-\Lambda^* : [0,1] \to \mathbb{R}$ *as the pointwise limit of* $-(\Lambda^\epsilon)^*$ *as* $\epsilon \to 0_+$*:*

$$-\Lambda^*(\theta) := \lim_{\epsilon \to 0_+} -(\Lambda^\epsilon)^*(\theta), \quad \text{for } \theta \in [0,1]. \tag{4.17}$$

Then $-\Lambda^*$ *is a continuous, concave function on* $[0,1]$*.*

Proof. From the definition of a typical set (4.3), it is easy to see that for $\epsilon_1 < \epsilon_2$, the corresponding typical sets satisfy $\mathcal{T}_n^{\epsilon_1} \subseteq \mathcal{T}_n^{\epsilon_2}$, for all $n \geq 1$. Using the monotonicity of intrinsic volumes with respect to inclusion and Theorem 4.1.3, we have $-(\Lambda^{\epsilon_1})^* \leq -(\Lambda^{\epsilon_2})^*$. Thus, for each $\theta \in [0,1]$, the value of $-(\Lambda^\epsilon)^*(\theta)$ monotonically decreases as $\epsilon \to 0_+$. To ensure that the quantity does not tend to $-\infty$ and establish a pointwise convergence result, we first provide a lower bound. From Lemma 2.2.3, we have

$$-(\Lambda^\epsilon)^*(0) \geq -(\Psi^\epsilon)^* = 0, \qquad \text{and} \tag{4.18}$$

$$-(\Lambda^\epsilon)^*(1) \geq -(\Psi^\epsilon)^*(1) \geq h(X) - \epsilon \tag{4.19}$$

By the concavity of $-(\Lambda^\epsilon)^*$, we obtain the linear lower bound $-(\Lambda^\epsilon)^*(\theta) \geq \theta(h(X)-\epsilon)$, for all $\theta \in [0,1]$. Thus, as $\epsilon \to 0_+$, the value of $-(\Lambda^\epsilon)^*$ may be uniformly lower-bounded. We then use the following lemma:

Lemma 4.2.2 (Proof in Appendix D.2.1)**.** *Let* $\{f_n\}$ *be a sequence of continuous, concave functions on* $[a,b]$*, converging pointwise and in a monotonically decreasing manner to a function* f*. Then* f *is a continuous, concave function on* $[a,b]$*.*

A simple application of Lemma 4.2.2 concludes the proof. □

Theorem 4.1.3 provides convergence in the large deviations sense. However, such convergence does not necessarily imply pointwise convergence, as desired in equation (4.6). Furthermore, as evidenced by Remark 2.2.5, the value of $-\Lambda^*$ at 0 and 1 is unknown; we only know that $-\Lambda^*$ is lower-bounded by 0 and $h(X)$ at these points. Since we want a smooth extension of the differential entropy function over $[0,1]$, we would like to show $-\Lambda^*(1) = h(X)$ and $-\Lambda^*(0) = 0$, and that the function h_θ from equation (4.6) is simply $-\Lambda^*$. We address these points in the following theorem:

Theorem 4.2.3. *The function* $-\Lambda^*$ *satisfies*

$$\lim_{\epsilon \to 0_+} \lim_{n \to \infty} \frac{1}{n} \log \mu_n^\epsilon(\lfloor n\theta \rfloor) = -\Lambda^*(\theta), \qquad \forall \theta \in [0,1], \tag{4.20}$$

where $\mu_n^\epsilon(\lfloor n\theta \rfloor)$ *is the* $\lfloor n\theta \rfloor^{th}$ *intrinsic volume of* $\mathcal{T}_n^\epsilon$*. In particular,* $-\Lambda^*(0) = 0$ *and* $-\Lambda^*(1) = h(X)$*.*

Proof. The proof consists of three separate parts based on the value of θ: (a) $\theta = 0$, (b) $\theta \in (0,1)$, and (c) $\theta = 1$.

Proof of (a): Let $\epsilon > 0$ be fixed. We will show that $-(\Lambda^\epsilon)^*(0) = 0$. This implies that $-\Lambda^*(0)$, which is the pointwise limit of $-(\Lambda^\epsilon)^*$ as $\epsilon \to 0_+$, also equals 0.

Since $\Phi(x) \to \pm\infty$ as $|x| \to \pm\infty$, we may find constants $c_1 > 0$ and c_2 such that

$$\Phi(x) \geq c_1|x| + c_2, \quad \text{for all} \quad x \in \mathbb{R}. \tag{4.21}$$

It is easy to check that for $A = \frac{h(X)+\epsilon-c_2}{c_1}$, the sequence of regular crosspolytopes $\{C_n\}_{n=1}^{\infty}$ defined by

$$C_n := \{x^n \in \mathbb{R}^n \mid \sum_{i=1}^{n} |x_i| \leq An\},$$

satisfies the containment $\mathcal{T}_n^\epsilon \subseteq C_n$, for $n \geq 1$.

Furthermore, the sequence $\{C_n\}_{n\geq 1}$ satisfies the same inclusion relation given in Lemma 4.1.1 for $\{\mathcal{T}_n^\epsilon\}$:

$$C_m \times C_n \subseteq C_{m+n}, \quad \forall m, n \geq 1.$$

Following a similar sequence of steps as for $\{\mathcal{T}_n^\epsilon\}$, one may check that the sequence of intrinsic volumes of $\{C_n\}$ converges in the large deviation sense to a function $-\chi^* : [0,1] \to \mathbb{R}$. It is possible to use the closed-form expression for intrinsic volumes of C_n [2] to show that $-\chi^*(0) = 0$, which we show in the following lemma:

Lemma 4.2.4 (Proof in Appendix D.2.2)**.** *The function $-\chi^*$, which is the $\mathcal{G}$-function of $\{C_n\}$ is continuous at 0; i.e. $-\chi^*(0) = 0$.*

The inclusion $\mathcal{T}_n^\epsilon \subseteq C_n$ leads to the inequality $-(\Lambda^\epsilon)^* \leq -\chi^*$. At $\theta = 0$, this yields $-(\Lambda^\epsilon)^*(0) \leq -\chi^*(0) = 0$. Combined with the lower bound $-(\Lambda^\epsilon)^* \geq 0$ from Lemma 2.2.3, we have $-(\Lambda^\epsilon)^*(0) = 0$.

Proof of (b): Let $\epsilon > 0$ be fixed. Let the intrinsic volumes of $\mathcal{T}_n^\epsilon$ be given by $\{\mu_n^\epsilon(0), \dots, \mu_n^\epsilon(n)\}$. We define a function $a_n : [0,1] \to \mathbb{R}$ by linearly interpolating the values at $a_n(j/n)$ for $0 \leq j \leq n$, where $a_n(j/n)$ is given by

$$a_n\left(\frac{j}{n}\right) = \frac{1}{n}\log\mu_n^\epsilon(j), \quad \text{for} \quad 0 \leq j \leq n. \tag{4.22}$$

By exactly the same techniques as in Lemma 3.8.6, we may show that for $\theta \in (0,1)$ the functions $\{a_n\}$ converge pointwise to $-(\Lambda^\epsilon)^*(\theta)$. This is shown by using the fact that intrinsic volumes form a log-concave sequence [23], in conjunction with the large deviations convergence from Theorem 4.1.3. Thus, for $\theta \in (0,1)$, equation (4.22) implies

$$\lim_{n\to\infty} \frac{1}{n}\log\mu_n^\epsilon(\lfloor n\theta \rfloor) = \lim_{n\to\infty} a_n(\theta) = -(\Lambda^\epsilon)^*(\theta). \tag{4.23}$$

Taking the limit as $\epsilon \to 0_+$, we obtain

$$\lim_{\epsilon\to 0_+}\lim_{n\to\infty} \frac{1}{n}\log\mu_n^\epsilon(\lfloor n\theta \rfloor) = -\Lambda^*(\theta). \tag{4.24}$$

Proof of (c): Proving $-\Lambda^*(1) = h(X)$ is more challenging than the previous cases. Log-concavity and large deviations type convergence alone are insufficient to pin down the value of $-\Lambda^*(1)$, as illustrated by remark 2.2.5. We use the following inequality for intrinsic volumes, proved in [4]:

Theorem 4.2.5. *Let $K \in \mathcal{C}_n$ and let the intrinsic volumes of K be denoted by $\{V_0(K), \dots, V_n(K)\}$. Let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Let $K|e_i^\perp$ be the set obtained by orthogonally projecting K on the $(n-1)$-dimensional subspace spanned by $\{e_1, \dots, e_{i-1}, e_{i+1}, e_n\}$. The following inequality holds:*

$$V_m(K) \le \frac{1}{n-m} \sum_{i=1}^n V_m(K \mid e_i^\perp), \tag{4.25}$$

provided the intrinsic volumes of $K|e_j^\perp$ satisfy the condition

$$(*) : V_m(K \mid e_j^\perp) \le \frac{1}{n-m} \sum_{i=1}^n V_m(K \mid e_i^\perp), \ \forall j \le n. \tag{4.26}$$

Ignoring $(*)$ for the moment, we iterate inequality (4.25) by applying it to each of the sets $K \mid e_i^\perp$, and then applying it again to the sets $K \mid \{e_i, e_j\}^\perp$ and continuing in a similar way, to arrive at the inequality

$$V_m(K) \le \sum_{1 \le i_1 < \cdots < i_{n-m} \le n} V_m(K \mid \{e_{i_1}, \dots, e_{i_{n-m}}\}^\perp). \tag{4.27}$$

We claim that condition $(*)$ is satisfied by all sets of the form $\mathcal{T}_n^\epsilon \mid \{e_{i_1}, \dots, e_{i_r}\}^\perp$. Without loss of generality, consider $A := \mathcal{T}_n^\epsilon \mid \{e_1, \dots, e_r\}^\perp$. If $M = \max_x p_X(x)$, the set $A \mid e_{r+1}^\perp$ consists of $(x_{r+2}, \dots, x_n) \in \mathbb{R}^{n-r-1}$ such that

$$\prod_{j=r+2}^n p_X(x_j) \ge \frac{e^{-n(h(X)+\epsilon)}}{M^{r+1}}. \tag{4.28}$$

From this expression, we observe that for any $j > 1$, the set $A \mid e_{r+j}^\perp$ has a similar description and is simply a rotation of the set $A \mid e_{r+1}^\perp$. Since intrinsic volumes are invariant under rotations, we conclude that the sets $\{A \mid e_{r+j}^\perp\}$ possess the same m^{th} intrinsic volume for all $1 \le j \le n-r$. If the value of the m^{th} intrinsic volumes equals $a > 0$, condition $(*)$ for set A is $a \le \frac{a(n-r)}{n-r-m}$, which holds trivially. Note also that when $r = n-m-1$, the sets $A \mid e_{r+j}^\perp$ are m-dimensional, so using inequality (4.28), we may conclude that

$$V_m(A \mid e_{r+j}^\perp) = \mathrm{Vol}(A \mid e_{r+j}^\perp) \le e^{n(h(X)+\epsilon)} M^{n-m}. \tag{4.29}$$

For $\theta \in (0,1)$, choose $m = \lfloor n\theta \rfloor$ and $K = \mathcal{T}_n^\epsilon$. Substituting in inequality (4.27) and using inequality (4.29), we obtain

$$\mu_n^\epsilon(\lfloor n\theta \rfloor) \le \binom{n}{\lfloor n\theta \rfloor} e^{n(h(X)+\epsilon)} M^{n-\lfloor n\theta \rfloor}.$$

Taking logarithms of both sides and dividing by n, we obtain

$$\frac{1}{n}\log \mu_n^{\epsilon}(\lfloor n\theta\rfloor) \le \frac{1}{n}\log\left(\binom{n}{\lfloor n\theta\rfloor} e^{n(h(X)+\epsilon)} M^{n-\lfloor n\theta\rfloor}\right).$$

Taking the limit as $n \to \infty$, and using part (b), we obtain

$$-\left(\Lambda^{\epsilon}\right)^{*}(\theta) \le H(\theta) + h(X) + \epsilon + (1-\theta)\log M, \tag{4.30}$$

where $H(\theta) = -\theta\log\theta - (1-\theta)\log(1-\theta)$ is the binary entropy function. We now take the limit as $\epsilon \to 0_+$, and use Theorem 4.2.1 to obtain

$$-\Lambda^*(\theta) \le H(\theta) + h(X) + (1-\theta)\log M. \tag{4.31}$$

Taking the limit $\theta \to 1$ and using continuity of $-\Lambda^*$ from Theorem 4.2.1, we have $-\Lambda^*(1) \le h(X)$. Combined with the lower bound from Lemma 2.2.3, which asserts that $-\Lambda^*(1) \ge h(X)$, we may conclude $-\Lambda^*(1) = h(X)$. This completes the proof. □

4.3 Alternate definitions of typical sets

For a log-concave random variable X, we defined its one-sided ϵ-typical set as

$$\mathcal{T}_n^{\epsilon} = \{x^n \in \mathbb{R}^n \mid p_{X^n}(x^n) \ge \exp\left(-n(h(X)+\epsilon)\right)\}.$$

We did this because of two reasons. Firstly, a one-sided typical set is larger that the traditional two-sided set, but this difference is negligibly small. Thus a one-sided typical set still satisfies the property that its growth rate is approximately the entropy of X, which was what we desired. Secondly, a one-sided typical set is convex . This enabled us to use convex geometric concepts such as intrinsic volumes in relation to these sets. However, there are alternative definitions of typical sets which achieve both these purposes, but are not amenable to the theory we have developed because they are not super-convolutive. A simple example of such an alternate definition may be

$$\mathcal{T}_n = \underset{A\in\mathcal{C}_n, P(A)\ge 0.99}{\arg\min} |A| \tag{4.32}$$

One may check for log-concave random variables, the above definition does indeed yield convex sets whose volume growth rate is the entropy of X. The constant 0.99 may be replaced by any other constant between 0 and 1 to achieve the same result. It is natural to examine if the $\mathcal{G}$-function of this alternate definition exists, and whether it equals $-\Lambda^*$. In this section, we consider a more general class of sequences of convex typical sets and show that the $\mathcal{G}$-function of all the sequences is $-\Lambda^*$.

Theorem 4.3.1. *Let X be a non-uniform log-concave random variable. For $\epsilon > 0$, define the following sequences of sets*

$$\overline{\mathcal{T}}_n^\epsilon = \{x^n \mid p_{X^n}(x^n) \geq \exp(-n(h(X)+\epsilon))\} \tag{4.33}$$
$$\underline{\mathcal{T}}_n^\epsilon = \{x^n \mid p_{X^n}(x^n) \geq \exp(-n(h(X)-\epsilon))\}. \tag{4.34}$$

Let $\{T_n\}$ be any sequence of compact convex sets. Suppose that for any $\epsilon > 0$ there exists an $N(\epsilon)$ such that for all $n > N(\epsilon)$ the following inclusion holds:

$$\underline{\mathcal{T}}_n^\epsilon \subseteq \mathcal{T}_n \subseteq \overline{\mathcal{T}}_n^\epsilon. \tag{4.35}$$

Then the $\mathcal{G}$-function of $\{T_n\}$ equals $-\Lambda^*$.

Remark 4.3.2. We exclude uniform random variables because $\underline{\mathcal{T}}_n^\epsilon = \phi$ for all n and all $\epsilon > 0$.

Proof. Note that it is enough to show that the $\mathcal{G}$-function of $\{\underline{\mathcal{T}}_n^\epsilon\}$ exists for all small enough ϵ, and tends to $-\Lambda^*$ as $\epsilon \to 0$. Using the same proof idea as in Lemma 4.1.1, we see that$\{\underline{\mathcal{T}}_n^\epsilon\}$ is a super-convolutive sequence. Using the non-uniformity of X, we observe that for all small enough ϵ the set $\underline{\mathcal{T}}_n^\epsilon \neq \phi$ for all n. Furthermore, noting that $\underline{\mathcal{T}}_n^\epsilon \subseteq \overline{\mathcal{T}}_n^\epsilon$ we see that the intrinsic volumes of $\{\underline{\mathcal{T}}_n^\epsilon\}$ satisfy all the required convergence properties as in Lemma 4.1.2. Thus, the $\mathcal{G}$-function of $\{\underline{\mathcal{T}}_n^\epsilon\}$ exists, although it may not be continuous at 1 (continuity at 0 follows since it is bounded above by $-\Lambda^*$.)

We will now show that $\mathcal{G}$-functions of $\{\underline{\mathcal{T}}_n^\epsilon\}$ and $\{\overline{\mathcal{T}}_n^\epsilon\}$, denoted by $\underline{\mathcal{G}}^\epsilon$ and $\overline{\mathcal{G}}^\epsilon$, cannot differ by too much. Our strategy is to show that it is possible to "bloat" $\{\underline{\mathcal{T}}_n^\epsilon\}$ by a small fraction, so that the bloated set will contain $\{\overline{\mathcal{T}}_n^\epsilon\}$.

Let $p_X(x) = \exp(-U(x))$, and assume without loss of generality that U achieves its minimum at 0. Note that $h(X) = \int U(x)\exp(U(x))dx$, and thus $h(X) \geq U(0)$. This inequality is strict when X is not uniform. Assume $\epsilon < h(X) - U(0)$. In this proof we assume that U is differentiable and is supported on $\mathbb{R}$, although the proof can be easily adapted to the case when it is not. The two sequences of sets may be equivalently described as

$$\overline{\mathcal{T}}_n^\epsilon = \{x^n \mid \sum U(x_i) \leq n(h(X)+\epsilon)\} \tag{4.36}$$
$$\underline{\mathcal{T}}_n^\epsilon = \{x^n \mid \sum U(x_i) \leq n(h(X)-\epsilon)\}. \tag{4.37}$$

For $\alpha > 0$, consider the map that takes $x^n \to (1+\alpha)x^n$. Let x^n be a point on the boundary of $\underline{\mathcal{T}}_n^\epsilon$ satisfying $\sum U(x_i) = n(h(X)-\epsilon)$. We have the inequalities

$$\begin{aligned}\sum U(x_i(1+\alpha)) &\geq \sum U(x_i) + \sum \alpha x_i U'(x_i)\\ &\geq n(h(X)-\epsilon) + \alpha \sum \big(U(x_i) - U(0)\big)\\ &\geq n(h(X)-\epsilon) + n\alpha(h(X)-\epsilon-U(0))\end{aligned}$$

Now for a choice of $\alpha = \frac{2\epsilon}{h(X)-\epsilon-U(0)}$, we will have

$$\sum U(x_i(1+\alpha)) \geq n(h(X)+\epsilon),$$

that is, $x^n(1+\alpha) \notin \overline{\mathcal{T}}_n^\epsilon$. Note that $\alpha \to 0$ as $\epsilon \to 0$. Thus, for this choice of α we must have

$$\overline{\mathcal{T}}_n^\epsilon \subseteq (1+\alpha)\underline{\mathcal{T}}_n^\epsilon.$$

This implies

$$\overline{\mathcal{G}}^\epsilon \leq \log(1+\alpha) + \underline{\mathcal{G}}^\epsilon.$$

We also have $\underline{\mathcal{G}}^\epsilon \leq \overline{\mathcal{G}}^\epsilon$, giving

$$\overline{\mathcal{G}}^\epsilon \leq \log(1+\alpha) + \underline{\mathcal{G}}^\epsilon \leq \log(1+\alpha) + \overline{\mathcal{G}}^\epsilon.$$

Since and $\lim_{\epsilon\to 0} \overline{\mathcal{G}}^\epsilon = -\Lambda^*$, we can take the limit as $\epsilon \to 0$ to conclude that

$$\lim_{\epsilon\to 0} \underline{\mathcal{G}}^\epsilon = -\Lambda^*.$$

This concludes the proof. □

4.4 Conclusion

The starting point of our work in this chapter was the relation between the volume of a typical set and the differential entropy of the associated distribution: Entropy is the exponential growth rate of the volume of typical sets. We subsequently generalized this relation beyond volumes to intrinsic volumes. Since intrinsic volumes are not defined for arbitrary sets, we considered log-concave distributions and defined their one-sided typical sets $\{\mathcal{T}_n^\epsilon\}$, which satisfy a crucial inclusion property:

$$(\mathbf{P}) : \mathcal{T}_m^\epsilon \times \mathcal{T}_n^\epsilon \subseteq \mathcal{T}_{m+n}^\epsilon,$$

hence implying that the intrinsic volumes of such sets are super-convolutive:

$$(\mathbf{P}_\mu) : \mu_m^\epsilon \star \mu_n^\epsilon \leq \mu_{m+n}^\epsilon.$$

We analyzed the convergence properties of such super-convolutive sequences. These convergence results in combination with certain geometric inequalities lead to our main result: There exists a continuous function $h_\theta : [0,1] \to \mathbb{R}$, such that for all log-concave random variables X, we have $h_0(X) = 0$, $h_1(X) = h(X)$, and $h_\theta(X)$ is the growth rate of the $n\theta^{\text{th}}$ intrinsic volume of the typical sets of X. Having shown the existence of $h_\theta(X)$, it remains to be seen what properties h_θ satisfies. We outline some future directions worth pursuing in Chapter 5.

Chapter 5

Discussion

Information theory and geometry closely parallel each other, and there is a huge potential for exchange of ideas amongst these fields. In this dissertation, we have shown how to reshape existing results in convex geometry so that they can be applied in an information-theoretic setting. In geometry, we often deal with sets in a fixed dimension whereas in information theory the dimension is not fixed and tends to infinity. We therefore studied certain geometric properties of *sequences of sets* instead of just one fixed set. The properties we chose to study, namely intrinsic volumes, are a fundamental component of convex geometry and emerge as natural candidates to analyze. Given a sequence of sets $\{K_n\}$, we described a recipe to find its $\mathcal{G}$-function; i.e. the growth rate of its intrinsic volumes:

1. Determine whether $\{K_n\}$ is sub or super-convolutive. As shown in Chapter 3, there is some freedom in this step as the sequence can also be "approximately" sub/super-convolutive.

2. Use convergence results for sub/sup-convolutive sequences to establish that for $\theta \in (0, 1)$, the normalized logarithm of the $n\theta^{\text{th}}$ intrinsic volume converges to a certain concave function $-\Lambda^*(\theta)$.

3. Show that $-\Lambda^*$ evaluated at 0 equals 0, and at 1 equals the growth rate of volume of $\{K_n\}$. There is as yet no standard procedure to achieve this step. In both the problems we considered, we had to use different techniques which relied on identifying key structural properties of the sequence being considered.

Once the existence of such a continuous $\mathcal{G}$-function is shown, then one may use a high dimensional version of Steiner's formula from convex geometry to find volumes of parallel bodies of $\{K_n\}$ in terms of its $\mathcal{G}$-function. This high dimensional Steiner's formula appears to be very useful in information theoretic applications, in particular to establish upper bounds on channel capacities. There are a number of open problems and future directions which are worth pursuing. We briefly describe some of them here.

5.1 Jump problem

This problem relates to step 3 in the recipe described above. In Chapter 3, to show that $-\Lambda^*$ evaluated at the endpoints 0 and 1 equalled the desired values 0 and $v(\sigma, \rho)$ our strategy involved showing that the sequence $\mathcal{S}_n(\sigma, \rho)$ is both sub and (approximately) super convolutive. For the typical sets considered in Chapter 4, such a strategy did not work. In order to resolve the value of $-\Lambda^*$ at the endpoints, we relied on two results. First, we showed that every typical set lies inside a regular crosspolytope and established that $-\Lambda^*(0) = 0$. Then we used a Loomis-Whitney type projection inequality for intrinsic volumes to show that $-\Lambda^*(1) = h(X)$. In both these problems, the methods used to rule out discontinuities, or jumps, in the $\mathcal{G}$-function required using highly problem-specific tools. We believe that there would be a considerable benefit in developing a procedure to get rid of these discontinuities which is general enough so as to be applicable to a large class of problems. In particular, the following conjecture regarding the $\mathcal{G}$-function of super-covolutive sets is still open:

Conjecture 1. *Let $\{K_n\}$ be a super-convolutive sequence such that the corresponding sequence of intrinsic volumes satisfies properties* $(\mathbf{A}), (\mathbf{B}), (\mathbf{C}), (\mathbf{D})$ *from Section 2.2. Then the $\mathcal{G}$-function of $\{K_n\}$ is continuous.*

So far in all the examples of super-convolutive sequences of sets that we have encountered, the $\mathcal{G}$-function has been continuous.

5.2 Intrinsic EPI problem

In Chapter 4, we showed the existence of $h_\theta(X)$ for a log-concave random variable X, which describes the geometry of the typical sets of X and leads to a generalization of its differential entropy. Having shown the existence, we would now like to discover what properties h_θ satisfies. As an example, we conjecture a version of the EPI inspired by the *complete* Brunn-Minkowski inequality for intrinsic volumes [30] which states that for $A, B \in \mathcal{C}_n$ and $m \geq 1$,

$$V_m(A)^{1/m} + V_m(B)^{1/m} \leq V_m(A \oplus B)^{1/m}. \tag{5.1}$$

A "complete" EPI could then be conjectured as:

Conjecture 2. *For real valued log-concave random variables X and Y, the following inequality holds:*

$$e^{\frac{2h_\theta(X)}{\theta}} + e^{\frac{2h_\theta(Y)}{\theta}} \leq e^{\frac{2h_\theta(X+Y)}{\theta}}, \tag{5.2}$$

where we recover the usual EPI for $\theta = 1$.

We believe a promising approach towards resolving this conjecture is that pursued in Szarek & Voiculescu [37]. In this paper, the authors provide an alternate proof of the EPI by proving a "restricted" version of Brunn-Minkowski inequality, stated as follows:

Theorem 5.2.1 (Restricted Brunn-Minkowski inequality)**.** *Let $A, B \subset \mathbb{R}^n$. For a set $\Theta \subseteq A \times B$, the restricted Minkowski sum (with respect to Θ) of A and B is the set*

$$A +_\Theta B = \{x + y \mid (x, y) \in \Theta\}.$$

For any $\epsilon > 0$, there exists a $\delta > 0$ such that the approximate Brunn-Minkowski inequality

$$|A +_\Theta B|^{2/n} \geq (1 - \epsilon) \left(|A|^{2/n} + |B|^{2/n}\right) \tag{5.3}$$

holds whenever the set Θ is large enough to satisfy

$$|\Theta| \geq (1 - \delta)^n |A \times B|. \tag{5.4}$$

We conjecture a version of restricted complete Brunn-Minkowski inequality, which implies Conjecture 2:

Conjecture 3. *Let $A, B \in \mathcal{C}_n$, and let $\Theta \subseteq A \times B$ be a convex set. The restricted Minkowski sum (with respect to Θ) of A and B is the convex set*

$$A +_\Theta B = \{x + y \mid (x, y) \in \Theta\}.$$

For any $\epsilon > 0$, there exists a $\delta > 0$ such that the approximate complete Brunn-Minkowski inequality

$$V_i(A +_\Theta B)^{2/i} \geq (1 - \epsilon) \left(V_i(A)^{2/i} + V_i(B)^{2/i}\right) \tag{5.5}$$

holds holds for all $i \geq 1$, whenever the set Θ is large enough to satisfy

$$|\Theta| \geq (1 - \delta)^n |A \times B|. \tag{5.6}$$

5.3 Subset problem

The subset problem is motivated by the typical set problem from Chapter 4. A very general definition of a convex typical sets can be as follows:

Definition 2. For a log-concave random variable X, a sequence of convex sets $\{\mathcal{T}_n\}$ is called typical if for all $\epsilon > 0$, there exists an $N(\epsilon)$ such that for all $n > N(\epsilon)$ we have

$$\mathcal{T}_n \subseteq \mathcal{T}_n^\epsilon,$$

where $\mathcal{T}_n^\epsilon$ is the usual one-sided typical set defined in equation (4.3). Furthermore, $\{\mathcal{T}_n\}$ should satisfy

$$\mathrm{liminf}_{n\to\infty} \mathbf{P}(\mathcal{T}_n) > 0.$$

It seems very likely that the $\mathcal{G}$-functions of such alternative definitions of typical sets also equals $h_\theta(X)$. Theorem 4.3.1 supports this belief. This would also provide additional evidence to the claim that h_θ is an intrinsic property of the distribution, and not something that depends on the way we define a typical set. It is as yet unclear how to best approach this problem.

5.4 Other future directions

Our generalization of entropy in Chapter 4 only works for log-concave distributions. Extending this definition for all distributions is an interesting direction. The main roadblock towards this is the lack of suitable intrinsic volume functionals for arbitrary sets. Although intrinsic volumes have been extended to some classes of nonconvex sets, such as star shaped bodies [19], a general extension for arbitrary bodies does not appear to exist.

Another possible direction is to redefine intrinsic entropy in a different way, which does not rely on typical sets at all. As mentioned earlier, Crofton's formula or Kubota's theorem related intrinsic volumes to the average size of lower dimensional projections or slices of convex bodies. In a similar vein, perhaps the i^{th} intrinsic entropy of a distribution on $\mathbb{R}^n$ could be defined using the average "size" of a random projection of a distribution. Here, a suitable substitute for size could be the entropy, or the entropy power. So the i^{th} intrinsic entropy of an $\mathbb{R}^n$ valued random variable X, not necessarily log-concave could be

$$V_i(X) \propto \int_{\mathcal{G}(n,i)} h(p_X \mid \mathcal{V}) d\mathcal{V}, \tag{5.7}$$

or

$$V_i(X) \propto \int_{\mathcal{G}(n,i)} \exp\left\{\frac{2}{i} h(p_X \mid \mathcal{V})\right\} d\mathcal{V}, \tag{5.8}$$

where $\mathcal{G}(n,i)$ is the collection of all i-dimensional subspaces of $\mathbb{R}^n$, and $h(p_X \mid V)$ is the entropy of the projection (marginal) of p_X on a subspace $\mathcal{V}$.

Random constant-dimensional marginals of log-concave distributions have been studied in the literature, for example in [21], where it is shown that most marginals are approximately Gaussian. However not much is known about random $n\theta$-dimensional marginals of log-concave distribution. While interesting in itself, a concentration result for such distributions can also assist in simplifying the alternative approaches to defining intrinsic entropy. Related to this, one may also study the volumes of random $n\theta$-dimensional projections of high-dimensional convex bodies. Some prior related work for constant dimensional projections of the n-dimensional cube can be found in [26]. The expected volume of a random $n\theta$-dimensional projection is known to be the $n\theta$-th intrinsic volume of the convex body, but not much is known about the distribution of this volume. It would be interesting to see if the distribution concentrates around the mean, and what rate it concentrates at as n becomes large.

Appendix A

Convergence results for convex functions

A.1 Pointwise and uniform convergence

Lemma A.1.1. *Let $\{f_n\}$ be a sequence of continuous convex functions which converge pointwise to a continuous function f on an interval $[a, b]$. Then f_n converge to f uniformly.*

Proof. Let $\epsilon > 0$. We'll show that there exists a large enough N such that for all $n > N$, $||f_n - f||_\infty < \epsilon$.

The function f is continuous on a compact set, and therefore is uniformly continuous. Choose a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon/10$ for $|x - y| < \delta$. Let M be such that $(b - a)/M < \delta$. We divide the interval $[a, b]$ into M intervals, whose endpoints are equidistant. We denote them by $a = \alpha_0 < \alpha_1 < \cdots < \alpha_M = b$. Since $f_n(\alpha_i) \to f(\alpha_i)$, there exists a N_i such that for all $n > N_i$, $|f_n(\alpha_i) - f(\alpha_i)| < \epsilon/10$. Choose $N = \max(M, N_0, \cdots, N_M)$.

Consider an $x \in (\alpha_i, \alpha_{i+1})$ for some $0 \leq i < M$, and let $n > N$. Using uniform continuity of f, we have

$$f(\alpha_i) - \epsilon/10 < f(x) < f(\alpha_i) + \epsilon/10. \tag{A.1}$$

Further, we also have

$$\begin{aligned} f_n(\alpha_i) &\leq f(\alpha_i) + \epsilon/10 \text{ , (by pointwise convergence at } \alpha_i) \\ f_n(\alpha_{i+1}) &\leq f(\alpha_{i+1}) + \epsilon/10 \text{ , (by pointwise convergence at } \alpha_{i+1}) \\ &\leq f(\alpha_i) + 2\epsilon/10 \text{ .(by uniform continuity of } f) \end{aligned}$$

Convexity of f_n implies

$$f_n(x) < \max(f_n(\alpha_i), f_n(\alpha_{i+1})) < f(\alpha_i) + 2\epsilon/10. \tag{A.2}$$

Combining part of equation (A.1) and equation (A.2), we obtain

$$f_n(x) - f(x) < 3\epsilon/10. \tag{A.3}$$

We'll now try to upper bound $f_n(x)$. First consider the case when $i \geq 1$. In this case we have

$$\alpha_{i-1} < \alpha_i < x < \alpha_{i+1}.$$

We write α_i as a linear combination of x and α_{i-1}, and use the convexity of f_n to arrive at

$$f_n(\alpha_i) \leq \frac{\alpha_i - \alpha_{i-1}}{x - \alpha_{i-1}} f_n(x) + \frac{x - \alpha_i}{x - \alpha_{i-1}} f_n(\alpha_{i-1}).$$

This implies

$$\frac{x - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} f_n(\alpha_i) - \frac{x - \alpha_i}{\alpha_i - \alpha_{i-1}} f_n(\alpha_{i-1}) \leq f_n(x).$$

Taking the infimum of the left side, we get

$$\inf_{x \in (\alpha_i, \alpha_{i+1})} \frac{x - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} f_n(\alpha_i) - \frac{x - \alpha_i}{\alpha_i - \alpha_{i-1}} f_n(\alpha_{i-1}) \leq f_n(x).$$

Note that since the LHS is linear in x, the infimum occurs at one of the endpoints of the interval, α_i or α_{i+1}. Substituting, we get

$$\begin{aligned} f_n(x) &\geq \min\left(f_n(\alpha_i), 2f_n(\alpha_i) - f_n(\alpha_{i-1})\right) \\ &\geq \min(f(\alpha_i) - \epsilon/10, 2(f(\alpha_i) - \epsilon/10) - f(\alpha_{i-1}) - \epsilon/10) \\ &\geq \min(f(\alpha_i) - \epsilon/10, 2f(\alpha_i) - f(\alpha_{i-1}) - 3\epsilon/10) \\ &\geq \min(f(\alpha_i) - \epsilon/10, 2f(\alpha_i) - f(\alpha_i) - \epsilon/10 - 3\epsilon/10) \\ &= f(\alpha_i) - 4\epsilon/10. \end{aligned} \tag{A.4}$$

Combining inequality (A.4) with a part of inequality (A.1), we have

$$f_n(x) - f(x) > -5\epsilon/10. \tag{A.5}$$

Combining (A.3) and (A.5) we conclude that for all $x \in (\alpha_1, \alpha_M)$, and for all $n > N$,

$$|f_n(x) - f(x)| < \epsilon/2. \tag{A.6}$$

Now let $x \in (\alpha_0, \alpha_1)$. We can establish inequality (A.3) for $x \in (\alpha_0, \alpha_1)$ using the same steps as above. We express α_1 as a linear combination of x and α_2 and follows the steps as above to establish (A.5) for $x \in (\alpha_0, \alpha_1)$. This shows that for all $x \in [a, b]$, $||f_n(x) - f(x)|| < \epsilon/2$ for all $n > N$, and concludes the proof. $\square$

A.2 Infimum over open sets

Lemma A.2.1. *Let $\{f_n\}$ be a sequence of continuous, convex functions on $[a,b]$, converging pointwise to f. Let $F \subseteq [a,b]$ be any relatively open set. Then*

$$\lim_n \big\{ \inf_{x\in F} f_n(x) \big\} = \inf_{x\in F} f(x).$$

Proof. The function f, being a pointwise limit of convex functions, is also convex on $[a,b]$. Since $f(0)$ and $f(1)$ are finite, it is possible to define a continuous convex function $\tilde{f}$ such that

$$\tilde{f}(x) = \begin{cases} f(x) & \text{for } \; a < x < b, \\ \lim_{x\to a} f(x) & \text{for } \; x = a, \\ \lim_{x\to b} f(x) & \text{for } \; x = b. \end{cases} \tag{A.7}$$

The function $\tilde{f}$ is continuous on a compact set, and therefore is uniformly continuous. Let $\epsilon > 0$ be given, and choose a $\delta > 0$ such that $|\tilde{f}(x) - \tilde{f}(y)| < \epsilon/10$ for $|x-y| < \delta$. We distinguish between two cases: (a) $\{a,b\} \cap F = \phi$, and (b) $\{a,b\} \cap F \neq \phi$. In case (a), we choose a $\delta' < \delta$ such that $F \cap [a, a+\delta'] = \phi$ and $F \cap [b-\delta', b] = \phi$. This means that F lies entirely in the set $[a+\delta', b-\delta']$. In case (b), we choose $\delta' < \delta$ small enough such that either $[a, a+\delta']$ and $[b-\delta', b]$ lie wholly in F, depending on whether $a \in F$ and $b \in F$ respectively. Let M be such that $(b-a)/M < \delta'$. We divide the interval $[a,b]$ into M intervals, whose endpoints are equidistant. We denote them by $a = \alpha_0 < \alpha_1 < \cdots < \alpha_M = b$. For $1 \leq i \leq M-1$, $\lim_n f_n(\alpha_i) \to \tilde{f}(\alpha_i)$, so there exists an N_i such that for all $n > N_i$, $|f_n(\alpha_i) - \tilde{f}(\alpha_i)| < \epsilon/10$. Choose $N = \max(N_1, \cdots, N_{M-1})$.

We express the interval $[0,1] = A \cup B \cup C$ where

$$\begin{aligned} A &= [0, \alpha_1] \\ B &= [\alpha_1, \alpha_{M-1}] \\ C &= [\alpha_{M-1}, \alpha_M]. \end{aligned}$$

Note that

$$\begin{aligned} \inf_{x\in F} f_n(x) &= \inf \left[\inf_{x\in F\cap A} f_n(x), \inf_{x\in F\cap B} f_n(x), \inf_{x\in F\cap C} f_n(x) \right] \\ &:= \inf[a_n, b_n, c_n]. \end{aligned}$$

If the intersection of F with either A, B or C is empty, we take the infimum to be $+\infty$.

Since f is upper semi-continuous, we have $f(0) \geq \tilde{f}(0)$ and $f(1) \geq \tilde{f}(1)$. Thus,

$$\begin{aligned} \inf_{x\in F} f(x) &= \inf \left[\inf_{x\in F\cap A} \tilde{f}(x), \inf_{x\in F\cap B} \tilde{f}(x), \inf_{x\in F\cap C} \tilde{f}(x) \right] \\ &:= \inf[a, b, c]. \end{aligned}$$

We will now show that for all $x \in B$,

$$|f_n(x) - \tilde{f}(x)| < \epsilon/2. \tag{A.8}$$

Consider an $x \in (\alpha_i, \alpha_{i+1})$ for some $1 \leq i \leq M-2$, and let $n > N$. Using uniform continuity of $\tilde{f}$, we have

$$\tilde{f}(\alpha_i) - \epsilon/10 < \tilde{f}(x) < \tilde{f}(\alpha_i) + \epsilon/10. \tag{A.9}$$

For $1 \leq i \leq M-2$, we also have

$$\begin{aligned} f_n(\alpha_i) &\leq \tilde{f}(\alpha_i) + \epsilon/10 \text{ , (by pointwise convergence)} \\ f_n(\alpha_{i+1}) &\leq \tilde{f}(\alpha_{i+1}) + \epsilon/10 \text{ , (by pointwise convergence)} \\ &\leq \tilde{f}(\alpha_i) + 2\epsilon/10 \text{ (by uniform continuity.)} \end{aligned}$$

Convexity of f_n implies

$$f_n(x) < \max(f_n(\alpha_i), f_n(\alpha_{i+1})) < \tilde{f}(\alpha_i) + 2\epsilon/10. \tag{A.10}$$

Combining part of equation (A.9) and equation (A.10), we obtain

$$f_n(x) - \tilde{f}(x) \leq 3\epsilon/10. \tag{A.11}$$

We'll now bound $f_n(x)$ from below. First consider the case when $i \geq 2$. In this case we have

$$\alpha_{i-1} < \alpha_i < x < \alpha_{i+1}.$$

We write α_i as a linear combination of x and α_{i-1}, and use the convexity of f_n to arrive at

$$f_n(\alpha_i) \leq \frac{\alpha_i - \alpha_{i-1}}{x - \alpha_{i-1}} f_n(x) + \frac{x - \alpha_i}{x - \alpha_{i-1}} f_n(\alpha_{i-1})$$

which implies

$$\frac{x - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} f_n(\alpha_i) - \frac{x - \alpha_i}{\alpha_i - \alpha_{i-1}} f_n(\alpha_{i-1}) \leq f_n(x),$$

leading to

$$\inf_{x \in (\alpha_i, \alpha_{i+1})} \frac{x - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} f_n(\alpha_i) - \frac{x - \alpha_i}{\alpha_i - \alpha_{i-1}} f_n(\alpha_{i-1}) \leq f_n(x).$$

Note that since the LHS is linear in x, the infimum occurs at one of the endpoints of the interval.

$$\begin{aligned} f_n(x) &\geq \min\left(f_n(\alpha_i), 2f_n(\alpha_i) - f_n(\alpha_{i-1})\right) \\ &\geq \min(\tilde{f}(\alpha_i) - \epsilon/10, 2(\tilde{f}(\alpha_i) - \epsilon/10) - \tilde{f}(\alpha_{i-1}) - \epsilon/10) \\ &\geq \min(\tilde{f}(\alpha_i) - \epsilon/10, 2\tilde{f}(\alpha_i) - \tilde{f}(\alpha_{i-1}) - 3\epsilon/10) \\ &\geq \min(\tilde{f}(\alpha_i) - \epsilon/10, 2\tilde{f}(\alpha_i) - \tilde{f}(\alpha_i) - \epsilon/10 - 3\epsilon/10) \\ &= \tilde{f}(\alpha_i) - 4\epsilon/10. \end{aligned} \tag{A.12}$$

Combining inequality (A.12) with a part of inequality (A.9), we have

$$f_n(x) - f(x) > -5\epsilon/10. \tag{A.13}$$

Combining (A.11) and (A.13) we conclude that for all $x \in (\alpha_2, \alpha_{M-1})$, and for all $n > N$,

$$|f_n(x) - f(x)| < \epsilon/2.$$

For $x \in (\alpha_1, \alpha_2)$, we can use a similar strategy as above. We bound $f_n(x)$ from above by expressing x as a convex combination of α_1 and α_2 and using the convexity of f_n. We then bound it from below by expressing α_2 as a linear combination of x and α_3 and using the convexity of f_n. This shows that for all $x \in [\alpha_1, \alpha_{M-1}]$, $|f_n(x) - f(x)| < \epsilon/2$ for all $n > N$, thus establishing equation (A.8). This implies that if $F \cap B \neq \phi$, then for all $n > N$

$$|b_n - b| \leq \epsilon/2. \tag{A.14}$$

Now suppose it were the case that $F \cap A \neq \phi$. This means that $[\alpha_0, \alpha_1] \subseteq F$, since we chose δ' to ensure this. We can lower bound f_n in the interval A using the convexity of f_n as follows. For an $x \in A$, we express α_1 as a linear combination of x and α_2 and obtain

$$f_n(\alpha_1) \leq f_n(x)\frac{\alpha_2 - \alpha_1}{\alpha_2 - x} + f_n(\alpha_2)\frac{\alpha_1 - x}{\alpha_2 - x}$$

which implies

$$\begin{aligned}
f_n(x) &\geq f_n(\alpha_1)\frac{\alpha_2 - x}{\alpha_2 - \alpha_1} - f_n(\alpha_2)\frac{\alpha_1 - x}{\alpha_2 - \alpha_1} \\
&\geq \inf_{x \in A} f_n(\alpha_1)\frac{\alpha_2 - x}{\alpha_2 - \alpha_1} - f_n(\alpha_2)\frac{\alpha_1 - x}{\alpha_2 - \alpha_1} \\
&\geq \min(f_n(\alpha_1), 2f_n(\alpha_1) - f_n(\alpha_2)) \\
&\geq \min\big(\tilde{f}(\alpha_1) - \epsilon/10, \\
&\qquad 2(\tilde{f}(\alpha_1) - \epsilon/10) - (\tilde{f}(\alpha_2) + \epsilon/10)\big) \\
&\geq \min(\tilde{f}(\alpha_1) - \epsilon/10, 2\tilde{f}(\alpha_1) - \tilde{f}(\alpha_2) - 3\epsilon/10) \\
&\geq \min(\tilde{f}(\alpha_1) - \epsilon/10, \\
&\qquad 2\tilde{f}(\alpha_1) - \tilde{f}(\alpha_1) - \epsilon/10 - 3\epsilon/10) \\
&= \tilde{f}(\alpha_1) - 4\epsilon/10.
\end{aligned}$$

Note that $|a - \tilde{f}(\alpha_1)| = |\inf_{x \in F \cap A} \tilde{f}(x) - f(\alpha_1)| < \epsilon/10$, thus giving us

$$f_n(x) \geq a - 5\epsilon/10 = a - \epsilon/2,$$

which implies

$$\inf_{x \in F \cap A} f_n(x) \geq a - \epsilon/2,$$

which leads to

$$a_n \geq a - \epsilon/2. \tag{A.15}$$

Furthermore, since $[\alpha_0, \alpha_1] \in F$, we have

$$\begin{aligned}\inf_{x\in F\cap A} f_n(x) = \inf_{x\in A} f_n(x) = a_n &\leq f_n(\alpha_1)\\ &\leq \tilde{f}(\alpha_1) + \epsilon/10\\ &\leq a + 2\epsilon/10. \end{aligned} \tag{A.16}$$

Combining inequalities (A.15) and (A.16), we conclude that for all $n > N$,

$$|a_n - a| \leq \epsilon/2. \tag{A.17}$$

Using a similar strategy as above, we can conclude that if $F \cap B \neq \phi$, then for all $n > N$,

$$|c_n - c| < \epsilon/2. \tag{A.18}$$

The inequalities (A.14), (A.17), and (A.18) imply that for all $n > N$,

$$\left|\inf_{x\in F} f_n(x) - \inf_{x\in F} f(x)\right| < \epsilon,$$

which completes the proof. $\square$

Appendix B

Proofs for Chapter 2

B.1 Proofs for Section 2.1

B.1.1 Proof of Lemma 2.1.1

1. The inequality (2.6) immediately gives that for all t, and all $n \geq 1$,

$$nG_1(t) \geq G_n(t), \quad \text{which implies} \quad g_1(t) \geq g_n(t). \tag{B.1}$$

Taking the limit in n, it follows that $\Lambda(t) \leq g_1(t)$ for all t.
For all n, the functions g_n are monotonically increasing, and for all t they satisfy

$$g_n(t) \geq \lim_{t \to -\infty} g_n(t) = \frac{1}{n} \log \mu_n(0). \tag{B.2}$$

In addition, we also know that

$$\inf_n \frac{1}{n} \log \mu_n(0) = \beta.$$

This gives us that

$$g_n(t) \geq \beta. \tag{B.3}$$

Taking the limit in n, we conclude that for all t,

$$\Lambda(t) \geq \beta. \tag{B.4}$$

For all n, we have the lower bound on g_n given by

$$\begin{aligned} g_n(t) &= \frac{1}{n} \log \sum_{j=0}^{n} \mu_n(j) e^{jt} && \text{(B.5)} \\ &\geq \frac{1}{n} \log \mu_n(n) e^{nt} && \text{(B.6)} \\ &= t + \frac{1}{n} \log \mu_n(n) && \text{(B.7)} \\ &\overset{(a)}{\geq} t + \alpha && \text{(B.8)} \end{aligned}$$

where (a) follows as

$$\inf_n \frac{1}{n} \log \mu_n(n) = \alpha.$$

Taking the limit in n, we conclude that

$$\Lambda(t) \geq t + \alpha. \tag{B.9}$$

Equations (B.4) and (B.9) establish

$$\Lambda(t) \geq \max(\beta, t + \alpha).$$

2. The functions $\{g_n\}$ are convex and monotonically increasing. Since Λ is the pointwise limit of these functions, Λ is also convex and monotonically increasing.

3. Note that the convex conjugates of the functions $g_1(t)$ and $\max(\beta, t+\alpha)$ are both supported on $[0, 1]$. Since Λ is trapped between these two functions, it is clear that Λ^* is also supported on $[0, 1]$.

B.2 Proof for Section 2.2

B.2.1 Proof of Lemma 2.2.1

1. Condition (2.30) implies that

$$nG_1(t) \leq G_n(t) \implies g_1(t) \leq g_n(t). \tag{B.10}$$

Taking the limit in n, we see that $g_1(t) \leq \Lambda(t)$.

For every n and every $t \geq 0$,

$$\begin{aligned} \frac{G_n(t)}{n} &= \frac{1}{n} \log \sum_{j=0}^{n} \mu_n(j) e^{jt} && \text{(B.11)} \\ &\leq \frac{1}{n} \log \left(\left(\sum_{j=0}^{n} \mu_n(j) \right) e^{nt} \right) && \text{(B.12)} \\ &= \frac{G_n(0)}{n} + t. && \text{(B.13)} \end{aligned}$$

Taking the limit in n and using assumption (**D**), we see that for $\Lambda(t) \leq t + \gamma$ for $t \geq 0$. Similarly, for $t \leq 0$

$$\frac{G_n(t)}{n} = \frac{1}{n} \log \sum_{j=0}^{n} \mu_n(j) e^{jt} \tag{B.14}$$

$$\leq \frac{1}{n} \log \left(\sum_{j=0}^{n} \mu_n(j) \right) \tag{B.15}$$

$$= \frac{G_n(0)}{n} \tag{B.16}$$

Taking the limit in n and using assumption (**D**), we see that $\Lambda(t) \leq \gamma$ for $t \leq 0$. This concludes the proof of part 1.

2. The functions $\{g_n\}$ are convex and monotonically increasing. Since Λ is the pointwise limit of these functions, Λ is also convex and monotonically increasing.

3. Note that the convex conjugates of the functions $g_1(t)$ and $\max(\gamma, t+\gamma)$ are both supported on $[0, 1]$. Since Λ is trapped between these two functions, it is clear that Λ^* is also supported on $[0, 1]$.

B.2.2 Proof of Lemma 2.2.3

We'll show that g_n^* converges pointwise to Λ^* on $(0, 1)$.
Fix an $x \in (0, 1)$, and define

$$\arg\max_t xt - g_n(t) := t_n.$$

Clearly, $g_n^*(x) = xt_n - g_n(t_n)$. Note that

$$g_n^*(x) \geq xt - g_n(t)\Big|_{t=0} \tag{B.17}$$

$$= -g_n(0) \tag{B.18}$$

$$\overset{(a)}{\geq} -\gamma \tag{B.19}$$

where (a) is because $\gamma = \sup_n g_n(0)$. If $t > \frac{\gamma - \log \mu_1(1)}{1-x}$, then we have

$$\begin{aligned} xt - g_n(t) &< xt - g_1(t) \\ &\leq xt - (t + \log \mu_1(1)) \\ &= -(1-x)t - \log \mu_1(1) \\ &< -(\gamma - \log \mu_1(1)) - \log \mu_1(1) \\ &= -\gamma. \end{aligned}$$

This gives us that $t_n \leq \frac{\gamma - \log \mu_1(1)}{1-x}$. Similarly, if $t < \frac{\log \mu_1(0) - \gamma}{x}$, then

$$\begin{aligned} xt - g_n(t) &< xt - g_1(t) && \text{(B.20)} \\ &\leq xt - \log \mu_1(0) && \text{(B.21)} \\ &< (\log \mu_1(0) - \gamma) - \log \mu_1(0) && \text{(B.22)} \\ &= -\gamma. && \text{(B.23)} \end{aligned}$$

This gives us that $t_n \geq \frac{\log \mu_1(0) - \gamma}{x}$. We can thus conclude that for all n,

$$t_n \in \left[\frac{\log \mu_1(0) - \gamma}{x}, \frac{\gamma - \log \mu_1(1)}{1-x} \right] := I_x. \tag{B.24}$$

Note that all we used to prove relation (B.24) is that g_n is trapped between g_1 and $\max(\gamma, t + \gamma)$. Since Λ also satisfies this, we have

$$\arg\max_t xt - \Lambda(t) \in I_x. \tag{B.25}$$

We now restrict our attention to the compact interval I_x. Let $\hat{g}_n$ be g_n restricted to I_x. The convex functions $\hat{g}_n$ converge pointwise to a continuous limit $\hat{\Lambda}$, where $\hat{\Lambda}$ is Λ restricted to I_x. This convergence must therefore be uniform, which implies convergence of $\hat{g}_n^*(x)$ to $\hat{\Lambda}^*(x)$. Furthermore, relation (B.24) implies $\hat{g}_n^*(x)$ equals $g_n^*(x)$, and relation (B.25) gives $\hat{\Lambda}^*(x)$ equals $\Lambda^*(x)$. Thus, $g_n^*(\cdot)$ converges pointwise to $\Lambda^*(\cdot)$ on $(0,1)$.

To get the inequality for $t = 0$ and $t = 1$, we note that both Ψ^* and Λ^* are convex on $[0,1]$ with Λ^* also being continuous on $[0,1]$. As both these function agree on $(0,1)$, it is immediate that $\Lambda^*(0) \leq \Psi^*(0) = -\beta$ and $\Lambda^*(1) \leq \Psi^*(1) = -\alpha$.

B.2.3 Super-convolutive sequence example

Define the generating function $A_n(x)$ of μ_n by

$$\begin{aligned} A_n(x) &= \sum_{i=0}^{n} \mu_n(i) x^i \\ &= (1 + \alpha x)^{n-1} + \epsilon x^n. && \text{(B.26)} \end{aligned}$$

For $m, n \geq 1$, the generating function of $\mu_m \star \mu_n$, is given by the product of their respective generating functions, $A_m(x) \times A_n(x)$. Without loss of generality, we assume

$m \leq n$. We'll write down this product explicitly as follows:

$$\begin{aligned} A_m(x)A_n(x) &= \sum_{j=0}^{m} \mu_m(j)x^j \sum_{i=0}^{n} \mu_n(i)x^i \\ &= \big((1+\alpha x)^{m-1} + \epsilon x^m\big)\big((1+\alpha x)^{n-1} + \epsilon x^n\big) \\ &= (1+\alpha x)^{m+n-2} + \epsilon x^m(1+\alpha x)^{n-1} + \epsilon x^n(1+\alpha x)^{m-1} + \epsilon^2 x^{m+n} \\ &=: \sum_{r=0}^{m+n} c_r x^r. \end{aligned}$$

The coefficients c_r are given by

$$c_r = \begin{cases} \binom{m+n-2}{r}\alpha^r & \text{for } 0 \leq r \leq m-1 \\ \binom{m+n-2}{r}\alpha^r + \epsilon\binom{n-1}{r-m}\alpha^{r-m} & \text{for } m \leq r \leq n-1 \\ \binom{m+n-2}{r}\alpha^r + \epsilon\binom{n-1}{r-m}\alpha^{r-m} + \epsilon\binom{m-1}{r-n}\alpha^{r-n} & \text{for } n \leq r \leq m+n-2 \\ \epsilon\alpha^{n-1} + \epsilon\alpha^{m-1} & \text{for } r = m+n-1 \\ \epsilon^2 & \text{for } r = m+n. \end{cases} \tag{B.27}$$

We shall now prove that the sequence μ_n is super convolutive, by showing that $\mu_{m+n}(r) \geq c_r$ for $0 \leq r \leq m+n$.

Case 1: $0 \leq r \leq m-1$

The following inequality trivially holds:

$$\binom{m+n-1}{r}\alpha^r \geq \binom{m+n-2}{r}\alpha^r. \tag{B.28}$$

Case 2: $m \leq r \leq n-1$

We need to show

$$\binom{m+n-1}{r}\alpha^r \geq \binom{m+n-2}{r}\alpha^r + \epsilon\binom{n-1}{r-m}\alpha^{r-m} \tag{B.29}$$

holds, which holds iff

$$\binom{m+n-2}{r-1}\alpha^r \geq \epsilon\binom{n-1}{r-m}\alpha^{r-m} \tag{B.30}$$

holds, which in turn holds iff

$$\binom{m+n-2}{r-1} \geq \epsilon\binom{n-1}{r-m}\alpha^{-m} \tag{B.31}$$

holds. Note that

$$\binom{m+n-2}{r-1} = \binom{(n-1)+(m-1)}{(r-m)+(m-1)} > \binom{n-1}{r-m} > \epsilon\binom{n-1}{r-m}\alpha^{-m}$$

by our assumptions on α and ϵ.

Case 3: $n \le r \le m+n-2$
We need to show that

$$\binom{m+n-1}{r}\alpha^r \ge \binom{m+n-2}{r}\alpha^r + \epsilon\binom{n-1}{r-m}\alpha^{r-m} + \epsilon\binom{m-1}{r-n}\alpha^{r-n} \tag{B.32}$$

holds, which holds iff

$$\binom{m+n-2}{r-1}\alpha^r \overset{?}{\ge} \epsilon\binom{n-1}{r-m}\alpha^{r-m} + \epsilon\binom{m-1}{r-n}\alpha^{r-n} \tag{B.33}$$

holds, which in turn holds iff

$$\binom{m+n-2}{r-1} \overset{?}{\ge} \epsilon\binom{n-1}{r-m}\alpha^{-m} + \epsilon\binom{m-1}{r-n}\alpha^{-n} \tag{B.34}$$

holds. Note that

$$\frac{1}{2}\binom{m+n-2}{r-1} = \frac{1}{2}\binom{(n-1)+(m-1)}{(r-m)+(m-1)} > \frac{1}{2}\binom{n-1}{r-m} > \epsilon\binom{n-1}{r-m}\alpha^{-m}$$

and

$$\frac{1}{2}\binom{m+n-2}{r-1} = \frac{1}{2}\binom{(m-1)+(n-1)}{(r-n)+(n-1)} > \frac{1}{2}\binom{m-1}{r-n} > \epsilon\binom{m-1}{r-n}\alpha^{-n}.$$

Adding these two gives us the required inequality.

Case 4: $r = m+n-1$
The following inequality is readily seen to hold, based on the choice $\alpha \ge 1$ and $\epsilon < \frac{1}{2}$:

$$\alpha^{m+n-1} \ge \epsilon\alpha^{n-1} + \epsilon\alpha^{m-1}, \tag{B.35}$$

Case 5: $r = m+n-1$ The following inequality hold trivially by the choice of $\epsilon < \frac{1}{2}$:

$$\epsilon \ge \epsilon^2, \tag{B.36}$$

As before, let $G_n(t) = \log\sum_{j=0}^{n}\mu_n(j)e^{jt}$. In our case, it equals

$$G_n(t) = \log\big((1+\alpha e^t)^{n-1} + \epsilon e^{nt}\big), \tag{B.37}$$

and

$$g_n(t) = \frac{1}{n}\log\big((1+\alpha e^t)^{n-1} + \epsilon e^{nt}\big). \tag{B.38}$$

Let $\Lambda(t) = \lim_{n\to\infty} g_n(t)$.

$$\begin{aligned}\Lambda(t) &= \lim_{n\to\infty} \frac{1}{n} \log\log\left((1+\alpha e^t)^{n-1} + \epsilon e^{nt}\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \log(1+\alpha e^t)^{n-1} + \frac{1}{n}\log\left(1 + \frac{\epsilon e^{nt}}{(1+\alpha e^t)^{n-1}}\right) \\ &= \log(1+\alpha e^t) + \lim_{n\to\infty} \frac{1}{n}\log\left(1 + \frac{\epsilon e^{nt}}{(1+\alpha e^t)^{n-1}}\right).\end{aligned}$$

We can use the bounds

$$\frac{\epsilon e^{nt}}{n\alpha^{n-1}e^{(n-1)t}} \leq \frac{\epsilon e^{nt}}{(1+\alpha e^t)^{n-1}} \leq \frac{\epsilon e^{nt}}{\alpha^{n-1}e^{(n-1)t}}, \tag{B.39}$$

which implies

$$\frac{\epsilon e^{t}}{n\alpha^{n-1}} \leq \frac{\epsilon e^{nt}}{(1+\alpha e^t)^{n-1}} \leq \frac{\epsilon e^{t}}{\alpha^{n-1}}. \tag{B.40}$$

Since $\alpha \geq 1$, the limit $\lim_{n\to\infty} \frac{1}{n}\log\left(1 + \frac{\epsilon e^{nt}}{(1+\alpha e^t)^{n-1}}\right)$ must equal 0.

Thus $\{g_n\}$ converge pointwise to $\Lambda(t) = \log(1+\alpha e^t)$. Let g_n^* be the Legendre-Fenchel transform of g_n, for $n \geq 1$. Note that g_n is a monotonically increasing convex function, with an asymptotic slope of 1 as $t \to \infty$. Thus, for all $n \geq 1$, we can compute $g_n^*(1)$ as follows:

$$\begin{aligned}g_n^*(1) &= \sup_t t - g_n(t) \\ &= \sup_t t - \frac{1}{n}\log\left((1+\alpha e^t)^{n-1} + \epsilon e^{nt}\right) \\ &= \lim_{t\to\infty} t - \frac{1}{n}\log\left((1+\alpha e^t)^{n-1} + \epsilon e^{nt}\right) \\ &= \frac{-\log\epsilon}{n}\end{aligned}$$

Thus,

$$\Psi^*(1) = \lim_{n\to\infty} g_n^*(1) = 0. \tag{B.41}$$

We can check that $\Lambda^*(1) = -\log\alpha \leq \Psi^*(1)$, with strict inequality if $\alpha > 1$.

Appendix C

Proofs for Chapter 3

C.1 Proofs for Section 3.4

C.1.1 Proof of Lemma 3.4.2

If we scale both σ and ρ by some $\alpha > 0$, by equation (3.4), $\mathcal{S}_n(\alpha\sigma, \alpha\rho)$ is a $\sqrt{\alpha}$-scaled version of $\mathcal{S}_n(\sigma, \rho)$. This means that $V_n(\alpha\sigma, \alpha\rho) = \alpha^{\frac{n}{2}} V_n(\sigma, \rho)$; i.e., $v(\alpha\sigma, \alpha\rho) = \log\sqrt{\alpha} + v(\sigma, \rho)$, which proves the lemma.

C.1.2 Proof of Lemma 3.4.3

Let $\sigma \in (0, \infty)$. Let $\epsilon > 0$ be given. We will show that there exists a $\delta^* > 0$ such that for all $\sigma' \in (\sigma - \delta^*, \sigma + \delta^*)$,

$$|v_1(\sigma') - v_1(\sigma)| < \epsilon.$$

Since v_1 is a non-decreasing function, it will be enough to show that

$$v_1(\sigma + \delta^*) - v_1(\sigma - \delta^*) < \epsilon.$$

Pick any $0 < \delta < \max(\sigma, \frac{1}{2})$. For $z \in (0, 1)$, let $\mathcal{S}_n(\sigma + \delta, 1) \times \sqrt{1 - z}$ denote the set $\mathcal{S}_n(\sigma+\delta, 1)$ scaled by $\sqrt{1 - z}$. Fix $z = 2\delta$. We will now show that $\mathcal{S}_n(\sigma+\delta, 1) \times \sqrt{1 - z} \subseteq S_n(\sigma - \delta, 1)$.

Any $(x_1, x_2, \ldots, x_n) \in \mathcal{S}_n(\sigma + \delta, 1)$ satisfies

$$\sum_{i=k+1}^{l} x_i^2 \leq (l - k) + \sigma + \delta \text{ for all } 0 \leq k < l \leq n. \tag{C.1}$$

Let the $(\hat{x}_1, \ldots, \hat{x}_n)$ be $(x_1, \ldots, x_n)$ scaled by $\sqrt{1 - z}$. If $(x_1, \ldots, x_n)$ happens to lie in $\mathcal{S}_n(\sigma - \delta, 1)$, then so does the scaled version $(\hat{x}_1, \ldots, \hat{x}_n)$. If $(x_1, \ldots, x_n) \in \mathcal{S}_n(\sigma + \delta, 1) \setminus$

$\mathcal{S}_n(\sigma - \delta, 1)$, then for each choice of $0 \le k < l \le n$ such that

$$(l-k) + \sigma + \delta \ge \sum_{i=k+1}^{l} x_i^2 > (l-k) + \sigma - \delta, \tag{C.2}$$

the point $(\hat{x}_1, \cdots, \hat{x}_n)$ satisfies

$$\sum_{i=k+1}^{l} \hat{x}_i^2 = \sum_{i=k+1}^{l} x_i^2 - z \sum_{i=k+1}^{l} x_i^2 \tag{C.3}$$

$$\le [(l-k) + \sigma + \delta] - z[(l-k) + \sigma - \delta] \tag{C.4}$$

$$\overset{(a)}{\le} [(l-k) + \sigma + \delta] - z \tag{C.5}$$

$$\overset{(b)}{=} (l-k) + \sigma - \delta, \tag{C.6}$$

where (a) follows since $l - k \ge 1$ and $\sigma - \delta > 0$, implying that $(l-k) + \sigma - \delta \ge 1$, and (b) follows by the choice $z = 2\delta$. Thus, the point $(\hat{x}_1, \ldots, \hat{x}_n)$ lies in the set $\mathcal{S}_n(\sigma - \delta, 1)$. The containment

$$\sqrt{1 - 2\delta} \times \mathcal{S}_n(\sigma + \delta, 1) \subseteq \mathcal{S}_n(\sigma - \delta, 1) \subseteq \mathcal{S}_n(\sigma + \delta, 1) \tag{C.7}$$

gives

$$\frac{1}{2}\log(1 - 2\delta) + v_1(\sigma + \delta) \le v_1(\sigma - \delta) \le v_1(\sigma + \delta). \tag{C.8}$$

Hence, we have

$$v_1(\sigma + \delta) - v_1(\sigma - \delta) \le -\frac{1}{2}\log(1 - 2\delta). \tag{C.9}$$

Picking δ^* small enough to satisfy

$$-\frac{1}{2}\log(1 - 2\delta^*) < \epsilon,$$

we establish continuity of $v_1(\sigma)$ in the open set $(0, \infty)$.

Now consider the case when $\sigma = 0$. We will show that there exists a $\delta^* > 0$ such that for all $\sigma' \in [0, \delta^*)$,

$$|v_1(\sigma') - v_1(0)| < \epsilon.$$

Since v_1 is a non-decreasing function, it will be enough to show that

$$v_1(\delta^*) - v_1(0) < \epsilon.$$

Pick any $\delta < 1$. Using the same strategy as before, we can show that $\mathcal{S}_n(\delta, 1) \times \sqrt{1 - \delta} \subseteq S_n(0, 1)$. This gives

$$v_1(\delta) + \frac{1}{2}\log(1 - \delta) \le v_1(0), \tag{C.10}$$

and thus

$$0 \le v_1(\delta) - v_1(0) \le -\frac{1}{2}\log(1-\delta). \tag{C.11}$$

Choosing δ^* small enough such that $-\frac{1}{2}\log(1-\delta) < \epsilon$, we establish continuity at $\sigma = 0$.

C.1.3 Proof of Lemma 3.4.4

For every n, define the function $V_n(\sigma) = \frac{\log \mathrm{Vol}(\mathcal{S}_n(\sigma,1))}{n}$. We'll first show that $V_n(\sigma)$ is concave. Define the set $\mathscr{S}_{n+1} \subseteq \mathbb{R}^{n+1}$ as follows:

$$\mathscr{S}_{n+1} = \{(x_1, \dots, x_n, \sigma_x) | (x_1, \dots, x_n) \in \mathcal{S}_n(\sigma_x, 1)\}. \tag{C.12}$$

We claim that $\mathscr{S}_{n+1}$ is convex. Let $\mathbf{x} = (x_1, \dots, x_n, \sigma_x)$ and $\mathbf{y} = (y_1, \dots, y_n, \sigma_y)$ be in $\mathscr{S}_{n+1}$. For $\lambda \in [0,1]$, consider the point $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$. For any $0 \le k < l \le n$, we have

$$\sum_{i=k+1}^{l} (\lambda x_i + (1-\lambda) y_i)^2 = \lambda^2 \sum_{i=k+1}^{l} x_i^2 + (1-\lambda)^2 \sum_{i=k+1}^{l} y_i^2 + 2\lambda(1-\lambda) \sum_{i=k+1}^{l} x_i y_i \tag{C.13}$$

$$\le \lambda^2 \sum_{i=k+1}^{l} x_i^2 + (1-\lambda)^2 \sum_{i=k+1}^{l} y_i^2 + \lambda(1-\lambda) \sum_{i=k+1}^{l} (x_i^2 + y_i^2) \tag{C.14}$$

$$= \lambda \sum_{i=k+1}^{l} x_i^2 + (1-\lambda) \sum_{i=k+1}^{l} y_i^2 \tag{C.15}$$

$$\le (\lambda\sigma_x + (1-\lambda)\sigma_y) + (l-k). \tag{C.16}$$

Thus, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathscr{S}_{n+1}$, which proves that $\mathscr{S}_{n+1}$ is a convex set.

Now the n-dimensional volume of the intersection of $\mathscr{S}_{n+1}$ with the hyperplane $\sigma_x = \sigma$ is simply the volume of $\mathcal{S}_n(\sigma, 1)$. Using the Brunn-Minkowski inequality [30], we see that $\mathrm{Vol}(\mathcal{S}_n(\sigma,1))^{\frac{1}{n}}$ is concave in σ, so the logarithm is also concave. This establishes the concavity of $V_n(\sigma)$.

To show that $v_1(\sigma)$ is concave, we simply note that it is the pointwise limit of the sequence of concave functions $\{V_n\}$.

C.1.4 Proof of Lemma 3.4.5

For $x^n \in \mathcal{A}_n(\sigma(n))$, the state at time n is nonnegative. Suppose that after time n, we impose a restriction that the power used per symbol cannot be more than $\frac{1}{2}$. This means that the battery will charge by at least $\frac{1}{2}$ at each timestep, and after $2\sigma(n)$ steps, the battery will be fully charged to $\sigma(n)$. Denote the set of all such $(n+2\sigma(n))$-length sequences obtained by this process as $\hat{\mathcal{A}}_n(\sigma(n))$. This set is contained in $\mathcal{S}_{n+2\sigma(n)}(\sigma(n), 1)$, and its volume is

$$\mathrm{Vol}(\mathcal{A}_n(\sigma(n))) \times (\sqrt{2})^{2\sigma(n)}.$$

The key point is to note the containment

$$\hat{\mathcal{A}}_n(\sigma(n)) \times \cdots \times \hat{\mathcal{A}}_n(\sigma(n)) \subset \mathcal{S}_{m(n+2\sigma(n))}(\sigma(n), 1) \ ,$$

for all $m \geq 1$, where there are m copies in the product on the left hand side. This holds because we ensure that the battery is fully charged to $\sigma(n)$ after each $(n+2\sigma(n))$-length block. Taking the limit in m and using Lemma 3.3.1, we see that

$$v_1(\sigma(n)) \geq \frac{1}{n+2\sigma(n)} \log \left(\mathrm{Vol}(\mathcal{A}_n(\sigma(n))) \times \sqrt{2}^{2\sigma(n)} \right).$$

Letting n tend to infinity and using conditions (3.48a) and (3.48b), we arrive at

$$\liminf_{n\to\infty} v_1(\sigma(n)) \geq \frac{1}{2} \log 2\pi e,$$

which proves the claim.

C.1.5 Proof of Lemma 3.4.6

The key to proving Lemma 3.4.6 is to examine the distribution of the burstiness $\sigma(X^n)$, when X^n is drawn from a uniform distribution on $\mathcal{A}_n$. Since a high-dimensional Gaussian closely approximates the uniform distribution on $\mathcal{A}_n$, it makes sense to look at the burstiness of X^n when each X_i is drawn independently from a standard normal distribution.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. standard normal random variables. Let $Y_i = X_i^2 - 1$, for $1 \leq i \leq n$. These Y_i are i.i.d. with zero mean and variance 2. Define $S_0 = 0$ and

$$S_m = \sum_{i=1}^{m} Y_i, \text{ for } 1 \leq m \leq n.$$

Define Σ_n, the burstiness of the sequence of X_i, by

$$\Sigma_n = \max_{0\leq k<l\leq n} Y_{k+1} + Y_{k+2} + \ldots + Y_l = \max_{0\leq k<l\leq n} (S_l - S_k).$$

The following inequality holds:

$$\Sigma_n \leq \max_{0\leq l\leq n} S_l - \min_{0\leq k\leq n} S_k := \tilde{\Sigma}_n. \tag{C.17}$$

Fix some $\alpha > 0$. Then

$$\begin{aligned}
\liminf_n P(\Sigma_n \leq \alpha\sqrt{n}) &\geq \liminf_n P(\tilde{\Sigma}_n \leq \alpha\sqrt{n}) \\
&= \liminf_n P(\max \frac{S_l}{\sqrt{2n}} - \min \frac{S_k}{\sqrt{2n}} \leq \frac{\alpha}{\sqrt{2}}) \\
&\overset{(a)}{=} P(\max_{0\leq t\leq 1} B_t - \min_{0\leq t\leq 1} B_t \leq \frac{\alpha}{\sqrt{2}}) \qquad \text{(C.18)} \\
&\geq P(\sup_{0\leq t\leq 1} |B_t| \leq \frac{\alpha}{2\sqrt{2}}),
\end{aligned}$$

where in equation (C.18), B_t is the standard Brownian motion and the equality in step (a) follows from Donsker's theorem [13]. We now choose α large enough so that

$$P(\sup_{0\le t\le 1} |B_t| \le \frac{\alpha}{2\sqrt{2}}) \ge \frac{3}{4}. \tag{C.19}$$

Since $\lim_n P(X^n \in \mathcal{A}_n) = 1/2$ by the central limit theorem for $Y_1, Y_2, \ldots$, we have

$$\liminf_n P(\mathcal{A}_n(\alpha\sqrt{n})) = \liminf_n P(\Sigma_n \le \alpha\sqrt{n}, X^n \in \mathcal{A}_n) \ge \frac{1}{4} \ ,$$

where $P(\mathcal{A}_n(\alpha\sqrt{n})) := P(X^n \in \mathcal{A}_n(\alpha\sqrt{n}))$. The volume of $\mathcal{A}_n(\alpha\sqrt{n}))$ is upper-bounded by the volume of $\mathcal{A}_n$. Furthermore, it is lower-bounded by the volume of a ball $\mathcal{B}_n$ centered at the origin, such that $P(X^n \in \mathcal{B}_n) =: P(\mathcal{B}_n) = P(\mathcal{A}_n(\alpha\sqrt{n}))$, since a Gaussian distribution decays radially. Using standard concentration bounds on the normal distribution [15], to satisfy $P(\mathcal{B}_n) \ge 1/4$, the radius of $\mathcal{B}_n$ must be $\sqrt{n}+o(\sqrt{n})$. Thus, $\lim_n \frac{\mathrm{Vol}(\mathcal{B}_n)}{n} = \frac{1}{2}\log 2\pi e$. Using $\mathrm{Vol}(\mathcal{B}_n) \le \mathrm{Vol}(\mathcal{A}_n(\alpha\sqrt{n})) \le \mathrm{Vol}(\mathcal{A}_n)$, we obtain $\lim_n \frac{\mathrm{Vol}(\mathcal{A}_n(\alpha\sqrt{n}))}{n} = \frac{1}{2}\log 2\pi e$. Thus, with $\sigma(n) = \alpha\sqrt{n}$, the pair $\mathcal{A}_n(\sigma(n))$ and $\sigma(n)$ satisfy both the conditions in Lemma 3.4.5, thereby proving Lemma 3.4.6.

C.2 Proofs for Section 3.5

Let $0 < \gamma < 1$. We define a new set $\mathcal{S}_{n,\gamma}(\sigma, 1)$ to be

$$\mathcal{S}_{n,\gamma}(\sigma,1) = \{x^n \in \mathcal{S}_n(\sigma,\rho) \mid x_i^2 \ge \gamma \text{ for every } 1 \le i \le n\}. \tag{C.20}$$

Using Fekete's Lemma, it is easy to establish that following limit exists:

$$\lim_{n\to\infty} \frac{\mathrm{Vol}(\mathcal{S}_{n,\gamma}(\sigma,1))}{n} := v_{1,\gamma}(\sigma). \tag{C.21}$$

Clearly, $v_{1,\gamma}(\sigma) \le v_1(\sigma)$ as $\mathcal{S}_{n,\gamma}(\sigma,1) \subseteq \mathcal{S}_n(\sigma,1)$. In Lemma C.2.1, we show that it is possible to choose a small enough value of γ such that $v_{1,\gamma}(\sigma)$ approximates $v_1(\sigma)$ as closely as desired.

Lemma C.2.1. *We have*

$$v_1\left(\frac{\sigma}{1-\eta}\right) + \frac{1}{2}\log(1-\eta) \le v_{1,\gamma}(\sigma) \le v_1(\sigma),$$

where $\eta = \gamma + 2(\sqrt{\sigma+1}\sqrt{\gamma})$.

Proof. Clearly, $v_{1,\gamma}(\sigma) \le v_1(\sigma)$, since $\mathcal{S}_{n,\gamma}(\sigma,1) \subseteq \mathcal{S}_n(\sigma,1)$.

Now let $x^n \in \mathcal{S}_n(\sigma, 1-\eta) \cap \mathbb{R}_+^n$. We claim that

$$(x_1+\sqrt{\gamma},\ldots,x_n+\sqrt{\gamma}) \in \mathcal{S}_{n,\gamma}(\sigma,1).$$

This would imply that a translated version of $\mathcal{S}_n(\sigma, 1-\eta) \cap \mathbb{R}_+^n$ lies inside $\mathcal{S}_{n,\gamma}(\sigma, 1) \cap \mathbb{R}_+^n$, which will give us a lower bound on the volume of the latter in terms of the former. Since each $(x_i + \sqrt{\gamma})^2 \geq \gamma$, the only condition we need to check is whether $(x_1 + \sqrt{\gamma}, \ldots, x_n + \sqrt{\gamma}) \in \mathcal{S}_n(\sigma, 1)$. For any $0 \leq k < l \leq n$, we have

$$\begin{aligned}
\sum_{i=k+1}^{l} (x_i + \sqrt{\gamma})^2 &= \sum_{i=k+1}^{l} x_i^2 + 2\sqrt{\gamma} \sum_{i=k+1}^{l} x_i + (l-k)\gamma && \text{(C.22)} \\
&\leq (l-k)(1-\eta) + \sigma + 2\sqrt{\gamma} \sum_{i=k+1}^{l} x_i + (l-k)\gamma && \text{(C.23)} \\
&\leq (l-k)(1-\eta+\gamma) + \sigma + 2\sqrt{\gamma} \sum_{i=k+1}^{l} \sqrt{\sigma+1} && \text{(C.24)} \\
&\leq (l-k)(1-\eta+\gamma+2\sqrt{\gamma}\sqrt{\sigma+1}) + \sigma. && \text{(C.25)} \\
&= (l-k) + \sigma && \text{(C.26)}
\end{aligned}$$

This gives us

$$\begin{aligned}
\mathrm{Vol}(\mathcal{S}_n(\sigma, 1-\eta) \leq \mathrm{Vol}(\mathcal{S}_{n,\gamma}(\sigma, 1)) \leq \mathrm{Vol}(\mathcal{S}_n(\sigma, 1)), && \text{(C.27)} \\
&& \text{(C.28)}
\end{aligned}$$

implying that

$$v_1\left(\frac{\sigma}{1-\eta}\right) + \frac{1}{2}\log(1-\eta) \leq v_{1,\gamma}(\sigma) \leq v_1(\sigma). \tag{C.29}$$

By the continuity of $v_1(\sigma)$, we see that choosing a γ (and consequently an η) small enough will give a value of $v_{1,\gamma}(\sigma)$ that is as close as desired to $v_1(\sigma)$. $\square$

Lemma C.2.1 ensures that a numerical method which can closely approximate $v_{1,\gamma}(\sigma)$ can also be used to closely approximate $v_1(\sigma)$ for small values of γ. Henceforth, we focus our attention on calculating $v_{1,\gamma}(\sigma)$. As noted in Section 3.5, we exploit the idea of battery state. Given $(x_1, \ldots, x_n) \in \mathcal{S}_{n,\gamma}(\sigma, 1)$, define

$$\phi_n = \begin{cases} \sigma_n & \text{if } \sigma_n < \sigma, \\ \sigma_{n-1} + 1 - x_n^2 & \text{if } \sigma_n = \sigma. \end{cases} \tag{C.30}$$

Setting $\phi_0 = \sigma$, equation (C.30) can also be written as

$$\phi_n = \begin{cases} \phi_{n-1} + 1 - x_n^2 & \text{if } \phi_{n-1} < \sigma, \\ \sigma + 1 - x_n^2 & \text{if } \phi_{n-1} \geq \sigma. \end{cases} \tag{C.31}$$

Define the function $\Phi_n : \mathcal{S}_{n,\gamma}(\sigma, 1) \to \mathbb{R}$ such that $\Phi_n(x_1, \ldots, x_n) = \phi_n$. Let λ_n be the Lebesgue measure restricted to $\mathcal{S}_{n,\gamma}(\sigma, 1)$. Let ν_n be the measure induced by Φ_n on $\mathbb{R}$. In Lemma C.2.2 below, we show the following:

Lemma C.2.2. *The measure ν_n is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$.*

Proof. We first calculate ν_1. Define

$$F_1(\phi_1) = \nu_1((-\infty, \phi_1]).$$

We have the relation $\phi_1 = \sigma + 1 - x_1^2$, where x_1 has the Lebesgue measure on $\mathcal{S}_{1,\gamma}(\sigma, 1)$: $[-\sqrt{\sigma+1}, -\sqrt{\gamma}] \cup [\sqrt{\sigma+1}, \sqrt{\gamma}]$. It is easy to see that

$$F_1(\phi) = \begin{cases} 0 & \text{for } \phi < 0 \\ 2(\sqrt{\sigma+1} - \sqrt{\sigma+1-\phi}) & \text{for } 0 \leq \phi \leq \sigma+1-\gamma \\ 2(\sqrt{\sigma+1} - \sqrt{\gamma}) & \text{for } \sigma+1-\gamma < \phi. \end{cases} \tag{C.32}$$

Observe that F_1, being Lipshitz, is an absolutely continuous function. This implies that the measure ν_1 is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and possesses a Radon-Nikodym derivative f_1, which equals the derivative of F_1 almost everywhere. We set f_1 as follows:

$$f_1(\phi) = \begin{cases} 0 & \text{for } \phi < 0 \\ \frac{1}{\sqrt{\sigma+1-\phi}} & \text{for } 0 \leq \phi \leq \sigma+1-\gamma \\ 0 & \text{for } \sigma+1-\gamma < \phi. \end{cases} \tag{C.33}$$

We note that f_1 is continuous and bounded on the closed interval $[0, \sigma+1-\gamma]$. Our proof now proceeds by induction. We assume that the measure ν_n admits a density f_n, which is continuous and bounded on the closed interval $[0, \sigma+1-\gamma]$, and prove that ν_{n+1} has a density f_{n+1} which is continuous and bounded on $[0, \sigma+1-\gamma]$.

Define

$$F_{n+1}(\phi) = \nu_{n+1}((-\infty, \phi]).$$

Since ν_n is supported on $[0, \sigma+1-\gamma]$, we can use the expression in (C.31) to conclude the same about ν_{n+1} and restrict our attention to $0 \leq \phi \leq \sigma+1-\gamma$. For ϕ in this range, we use relation (C.31) and express F_{n+1} in terms of f_n as follows:

$$\begin{aligned}
F_{n+1}(\phi) &= \int_{x=0}^{\phi-1+\gamma}\int_{t=\gamma}^{x+1}\frac{f_n(x)}{\sqrt{t}}dtdx \\
&+ \int_{x=\phi+1-\gamma}^{\sigma}\int_{t=x-(\phi-1)}^{x+1}\frac{f_n(x)}{\sqrt{t}}dtdx \\
&+ \int_{x=\sigma}^{\sigma+1-\gamma}\int_{t=\sigma-(\phi-1)}^{\sigma+1}\frac{f_n(x)}{\sqrt{t}}dtdx \qquad (C.34) \\
&= \int_{x=0}^{\phi-1+\gamma} 2f_n(x)[\sqrt{x+1}-\sqrt{\gamma}]dx \\
&+ \int_{x=\phi-1+\gamma}^{\sigma} 2f_n(x)[\sqrt{x+1}-\sqrt{x-(\phi-1)}]dx \\
&+ \int_{x=\sigma}^{\sigma+1-\gamma} 2f_n(x)[\sqrt{\sigma+1}-\sqrt{\sigma-(\phi-1)}]dx. \qquad (C.35)
\end{aligned}$$

From the induction assumption of continuity and boundedness of f_n, it is easy to check that F_{n+1} is Lipshitz and therefore absolutely continuous. This implies that ν_{n+1} permits a density, which is equal to the derivative of F_n almost everywhere. We can evaluate this density by differentiating F_{n+1} with respect to ϕ. This involves differentiating under the integral sign, and the conditions for doing so are seen to be satisfied because of the continuity and boundedness of f_n and the square root function. We then get

$$f_{n+1}(\phi) = \int_{\phi-1+\gamma}^{\sigma}\frac{f_n(x)}{\sqrt{x-(\phi-1)}}dx + \int_{\sigma}^{\sigma+1-\gamma}\frac{f_n(x)}{\sqrt{\sigma-(\phi-1)}}dx, \qquad (C.36)$$

which is supported on, and is bounded and continuous on, the interval $[0,\sigma+1-\gamma]$. □

Equation (C.36) in the proof of Lemma C.2.2 describes the evolution of f_n as the dimension n increases. Let $C([0,\sigma+1-\gamma]$ be the set of continuous functions defined on the interval $[0,\sigma+1-\gamma]$. Define the integral operator $A : C([0,\sigma+1-\gamma] \to C([0,\sigma+1-\gamma]$ as follows:

$$A(x,t) = \begin{cases} \frac{1}{\sqrt{x+1-t}} & \text{if } 0\le x<\sigma,\ 0\le t\le x+1-\gamma\ , \\ \frac{1}{\sqrt{\sigma+1-t}} & \text{if } \sigma\le x\le\sigma+1-\gamma,\ 0\le t\le\sigma+1-\gamma\ , \\ 0 & \text{otherwise.}\end{cases} \qquad (C.37)$$

We can express equation (C.36) in another form,

$$f_{n+1}(t) = \int A(x,t)f_n(x)dx. \qquad (C.38)$$

We denote this $f_{n+1} = A(f_n)$. Iterating this relation, we obtain

$$f_{n+1} = A^n f_1. \tag{C.39}$$

We make three crucial observations. Firstly, the kernel A is bounded and piecewise continuous with the discontinuities confined to a single curve $t = x + 1 - \gamma$. It is also immediate that the spectral radius of A, defined by

$$r(A) = \sup_{||f||=1} ||Af||$$

is such that $r(A) > 0$. We use Theorem 2.13 from Anselone [1] to obtain that such an operator A is compact. In addition, we can apply the Krein Rutman theorem from Schaefer [29] to establish that $r(A)$ is an eigenvalue with a positive eigenvector $u \in C([0, \sigma + 1 - \gamma] \setminus 0$.

Secondly, we have

$$\nu_n([0, \sigma + 1 - \gamma]) = \int_{x=0}^{\sigma+1-\gamma} f_n(x)dx = \mathrm{Vol}(\mathcal{S}_{n,\gamma}(\sigma, 1).$$

Thus we have

$$\begin{aligned}
v_{1,\gamma}(\sigma) &= \lim_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_{n,\gamma}(\sigma, 1)) && \text{(C.40)}\\
&= \lim_{n\to\infty} \frac{1}{n} \log \int_{x=0}^{\sigma+1-\gamma} f_n(x)dx && \text{(C.41)}\\
&= \lim_{n\to\infty} \frac{1}{n} \log \int_{x=0}^{\sigma+1-\gamma} A^{n-1} f_1(x)dx && \text{(C.42)}\\
&\overset{(a)}{=} r(A) && \text{(C.43)}
\end{aligned}$$

where (a) follows because the projection of f_1 in the direction of u is nonzero owing to the positivity of both these functions.

Thirdly, define a sequence of operators $\{A_n\}$ as discrete approximations of A as follows. Let $h_n = \frac{\sigma+1-\gamma}{n}$,

$$A_n f(t) = \sum_{j=0}^{n} A(jh_n, t) f(jh_n) h_n.$$

Using Theorem 2.13 from Anselone [1] once more, we conclude that the sequence of operators $\{A_n\}$ is collectively compact and that $||A_n|| \to ||A||$. We can now use existing numerical techniques to find $r(A_n)$, which will provide an approximation to $r(A)$. The spectral radius $r(A)$ equals $v_{1,\gamma}(\sigma)$, which closely approximates $v_1(\sigma)$, and validates the numerical procedure as described in Section 3.5.

C.3 Proofs for Section 3.7

C.3.1 Proof of Lemma 3.7.2

Denote $A_n = [-A, A]^n$, and $B_n = B_n(\sqrt{n\nu})$. Let $C_n = A_n \oplus B_n$. Note that for any $m, n \geq 1$

$$B_n(\sqrt{n\nu}) \times B_m(\sqrt{m\nu}) \subseteq B_{m+n}(\sqrt{(m+n)\nu}) \tag{C.44}$$

$$[-A, A]^n \times [-A, A]^m = [-A, A]^{m+n}. \tag{C.45}$$

It follows that

$$C_m \times C_n = (A_m \oplus B_m) \times (A_n \oplus B_n) \tag{C.46}$$

$$= (A_m \times A_n) \oplus (B_m \times B_n) \tag{C.47}$$

$$\subseteq A_{m+n} \oplus B_{m+n} \tag{C.48}$$

$$= C_{m+n}. \tag{C.49}$$

This implies

$$\mathrm{Vol}(C_{m+n}) \geq \mathrm{Vol}(C_m)\mathrm{Vol}(C_m), \tag{C.50}$$

which immediately implies existence of the limit $\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Vol}(C_n)$, which equals $\ell(\nu)$ as defined in equation (3.71). To show this limit is finite, we note that $A_n \subseteq B_n(\sqrt{nA^2})$. Thus $C_n \subseteq B_n(\sqrt{n}(\sqrt{\nu} + A))$, which gives

$$\ell(\nu) \leq \frac{1}{2} \log 2\pi e(\sqrt{\nu} + A)^2 < \infty.$$

C.3.2 Proof of Lemma 3.7.3

We have the trivial bounds

$$\frac{\mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu}))}{n+1} \leq e^{nf_n^\nu(\hat{\theta}_n)} \leq \mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})), \tag{C.51}$$

which implies

$$\frac{1}{n} \log \mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})) - \frac{\log(n+1)}{n} \leq f_n^\nu(\hat{\theta}_n) \leq \frac{1}{n} \log \mathrm{Vol}([-A, A]^n \oplus B_n(\sqrt{n\nu})). \tag{C.52}$$

Taking the limit in n and using Lemma 3.7.2 we see that

$$\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n) = \ell(\nu). \tag{C.53}$$

C.3.3 Proof of Lemma 3.7.4

We first prove pointwise convergence. Looking at equation (3.78), we see that all we need to prove is that for all $\theta \in [0,1]$,

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\frac{\Gamma(n+1)n^{n\theta/2}}{\Gamma(n(1-\theta)+1)\Gamma(n\theta+1)\Gamma(n\theta/2+1)}\right) = H(\theta) + \frac{\theta}{2}\log 2e - \frac{\theta}{2}\log\theta. \tag{C.54}$$

For $\theta = 0$, we can easily check the validity of this statement. Let $\theta > 0$. We use the approximation

$$\log\Gamma(z) = z\log z - z + \log\frac{z}{2\pi} + o(z).$$

$$\frac{1}{n}\log\left(\frac{\Gamma(n+1)n^{n\theta/2}}{\Gamma(n(1-\theta)+1)\Gamma(n\theta+1)\Gamma(n\theta/2+1)}\right) =$$
$$\frac{1}{n}\Bigg((n+1)\log\frac{n+1}{e} + \frac{n\theta}{2}\log n - (n(1-\theta)+1)\log\frac{n(1-\theta)+1}{e} - (n\theta+1)\log\frac{n\theta+1}{e}$$
$$-(n\theta/2+1)\log\frac{n\theta/2+1}{e} + o(n)\Bigg). \tag{C.55}$$

Using $(x+1)\log(x+1) = x\log x + o(x)$, we can simplify the above to get

$$\frac{1}{n}\left(n\log n + \frac{n\theta}{2}\log n - n\bar{\theta}\log n\bar{\theta} - n\theta\log n\theta - (n\theta/2)\log n\theta/2e + o(n)\right), \tag{C.56}$$
$$= \frac{1}{n}\left(nH(\theta) - (n\theta/2)\log(\theta/2e) + o(n)\right). \tag{C.57}$$

Taking the limit as $n \to \infty$, we establish equality (C.54).

To show uniform convergence, we first observe that the functions $f_n^\nu(\cdot)$ are concave. This concavity is immediately evident from the log-convexity of the Γ function and from equation (3.78). Therefore, $\{f_n^\nu\}$ are concave functions converging pointwise to a continuous functions f^ν on $[0,1]$. Uniform convergence now follows from Lemma A.1.1.

C.3.4 Proof of Lemma 3.7.5

By Lemma 3.7.4, the sequence of functions $\{f_n^\nu\}$ converges to f^ν uniformly. This uniform convergence implies that the family of functions $\{f_n^\nu\}$ is equicontinuous [28] (Section 10.1, Theorem 3, pg. 209). Let $\epsilon > 0$. Choose N large such that $|f_n^\nu(x) - f_n^\nu(y)| < \epsilon/2$ if $|x-y| < 1/N$. This implies that for all $n > N$,

$$\max_\theta f_n^\nu(\theta) \geq f_n^\nu(\hat{\theta}_n) > \max_\theta f_n^\nu(\theta) - \epsilon/2. \tag{C.58}$$

Using the uniform convergence of $\{f_n^\nu\}$, we choose M large enough such that $||f^\nu - f_n^\nu||_\infty < \epsilon/2$ for all $n > M$. Let $L = \max(M,N)$. For all $n > L$, we have

$$\max_\theta f^\nu(\theta) + \epsilon/2 > \max_\theta f_n^\nu(\theta) \geq f_n^\nu(\hat{\theta}_n) \geq \max_\theta f_n^\nu(\theta) - \epsilon/2 \geq \max_\theta f^\nu(\theta) - \epsilon,$$

and thus

$$|f_n^\nu(\hat{\theta}_n) - \max_\theta f^\nu(\theta)| < \epsilon.$$

This concludes the proof of equation (3.82). By Lemma 3.7.3, we immediately have the equality (3.83).

C.3.5 Proof of Theorem 3.7.7

Let $\epsilon > 0$. Let $\{X_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$ be n i.i.d copies of X and Z respectively. Let δ_n be given by

$$\delta_n := \mathbb{P}\left(X^n + Z^n \notin [-A, A]^n \oplus B_n(\sqrt{n(\nu+\epsilon)}\ )\right) \tag{C.59}$$

Denote

$$C_n := [-A, A]^n \oplus B_n(\sqrt{n(\nu+\epsilon)}\ ).$$

By the law of large numbers, the probability $\delta_n \to 0$.

Let $Y := X + Z$. We have

$$\begin{aligned} nh(Y) = h(Y^n) &= H(\delta_n) + (1-\delta_n)h(Y^n|Y^n \in C_n) + \delta_n h(Y^n|Y^n \notin C_n) && \text{(C.60)}\\ &\le H(\delta_n) + (1-\delta_n)\log \mathrm{Vol}(C_n) + \delta_n h(Y^n|Y^n \notin C_n). && \text{(C.61)} \end{aligned}$$

Let $\hat{Y}^n \sim p(Y^n|Y^n \notin C_n)$. We have following bound on Y^n

$$E[||Y^n||^2] \le n(\nu + A^2). \tag{C.62}$$

This translates to a bound on $\hat{Y}^n$

$$E[||\hat{Y}^n||^2] \le \frac{n(\nu+A^2)}{\delta_n}, \tag{C.63}$$

which implies

$$h(\hat{Y}^n) \le \frac{n}{2}\log\frac{2\pi e(\nu+A^2)}{\delta_n}. \tag{C.64}$$

Substituting in inequality (C.61),

$$h(Y^n) \le H(\delta_n) + (1-\delta_n)\log \mathrm{Vol}(C_n) + \delta_n \frac{n}{2}\log\frac{2\pi e(\nu+A^2)}{\delta_n} \tag{C.65}$$

which implies

$$h(Y) \le \frac{H(\delta_n)}{n} + (1-\delta_n)\frac{\log \mathrm{Vol}(C_n)}{n} + \frac{\delta_n}{2}\log\frac{2\pi e(\nu+A^2)}{\delta_n}. \tag{C.66}$$

Taking the limit in n, we get

$$h(Y) \le \ell(\nu + \epsilon). \tag{C.67}$$

As this holds for any choice of ϵ, we let ϵ tend to 0 and use the continuity from Theorem 3.7.1 to arrive at

$$h(Y) \le \ell(\nu). \tag{C.68}$$

C.4 Proofs for Section 3.8

C.4.1 Proof of Lemma 3.8.2

Let $x^n, y^n \in \mathcal{S}_n(\sigma,\rho)$ and let $z^n = \lambda x^n + (1-\lambda)y^n$. By Jensen's inequality we have for every $1 \leq i \leq n$,

$$z_i^2 \leq \lambda x_i^2 + (1-\lambda)y_i^2.$$

Since both x^n and y^n both satisfy (3.4), the above inequality gives us that z^n does so too; i.e., $z^n \in \mathcal{S}_n(\sigma,\rho)$.

C.4.2 Proof of Lemma 3.8.3

The sets $\{\mathcal{S}_n(\sigma,\rho)\}$ satisfy the containment

$$\mathcal{S}_{m+n} \subseteq \mathcal{S}_m \times \mathcal{S}_n \quad \text{for every } m,n \geq 1. \tag{C.69}$$

This implies that the family of intrinsic volumes $\{\mu_n(\cdot)\}_{n\geq 1}$, is *sub-convolutive*; i.e., it satisfies the following condition:

$$\mu_m \star \mu_n \geq \mu_{m+n} \text{ for every } m,n \geq 1. \tag{C.70}$$

Noting that $\mu_n(n)$ is the volume of $\mathcal{S}_n(\sigma,\rho)$, and $\mu_n(0)=1$ for all $\mathcal{S}_n$, we can check that the sequence $\{\mu_n(\cdot)\}$ satisfies the assumptions $(\mathbf{A})$, $(\mathbf{B})$ and $(\mathbf{C})$ detailed in Section 2.1; namely,

$$\begin{aligned}
&(\mathbf{A}) : \alpha := \lim_{n\to\infty} \frac{1}{n}\log \mu_n(n) \quad \text{ is finite.}\\
&(\mathbf{B}) : \beta := \lim_{n\to\infty} \frac{1}{n}\log \mu_n(0) \quad \text{ is finite.}\\
&(\mathbf{C}) : \text{For all } n,\ \mu_n(n) > 0, \mu_n(0) > 0.
\end{aligned}$$

Lemma 3.8.3 then follows from the results in Section 2.1, in particular Lemma 2.1.1.

C.4.3 Proof of Lemma 3.8.4

Note that the claims in points 1 and 2 immediately imply 3, since $f_n^\nu = a_n + b_n^\nu$.

We shall prove 2 first. The expression for $b_n^\nu(\theta)$ is given by

$$b_n^\nu(\theta) = \frac{1}{n}\log\frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)}(n\nu)^{n\theta/2} \quad \text{for} \quad \theta\in[0,1]. \tag{C.71}$$

Since the Gamma function is log-convex [3] (Exercise 3.52), we see that $b_n^\nu(\cdot)$ is a concave function.

To show 1, note that all we need to prove is that

$$a_n\left(\frac{j}{n}\right) \geq \frac{a_n\left(\frac{j-1}{n}\right) + a_n\left(\frac{j+1}{n}\right)}{2} \quad \text{for all} \quad 1 \leq j \leq n-1, \tag{C.72}$$

as a_n is a linear interpolation of the values at $\frac{j}{n}$. This is equivalent to proving

$$\mu_n(j)^2 \geq \mu_n(j-1)\mu_n(j+1) \quad \text{for all} \quad 1 \leq j \leq n-1. \tag{C.73}$$

This is an easy application of the Alexandrov-Fenchel inequalities for mixed volumes. For a proof we refer to McMullen [23], where in fact the author obtains

$$\mu_n(j)^2 \geq \frac{j+1}{j}\mu_n(j-1)\mu_n(j+1).$$

C.4.4 Proof of Lemma 3.8.5

As noted in Appendix C.4.2, the family of intrinsic volumes $\{\mu_n(\cdot)\}_{n\geq 1}$, is sub-convolutive and it satisfies the assumptions $(\mathbf{A}), (\mathbf{B})$, and $(\mathbf{C})$ detailed in Section 2.1. Part 1 of Lemma 3.8.5 is now an immediate consequence of Theorem 2.1.2.

To prove part 2, let $F \subseteq \mathbb{R}$ be an open set. We assume that $F \cap [0,1]$ is nonempty, since the otherwise the result is trivial. We will construct a new sequence of functions $\{\hat{\mu}_n\}$ such that $\mu_n \geq \hat{\mu}_n$ for all n; i.e., μ_n pointwise dominates $\hat{\mu}_n$ for all n. The large deviations lower bound for the sequence $\{\hat{\mu}_n\}$ will then serve as a large deviations lower bound for the sequence $\{\mu_n\}$.

For notational convenience, we write $\mathcal{S}_n$ for $\mathcal{S}_n(\sigma,\rho)$ in this proof. Fix an $a \geq 1$. Let $\gamma = \lceil \frac{\sigma}{\rho} \rceil$. Let

$$\hat{\mathcal{S}}_{a+\gamma} = \{x^{a+\gamma} \in \mathbb{R}^{a+\gamma} | x^a \in \mathcal{S}_a, x_{a+1}^{a+\gamma} = \mathbf{0}\}.$$

For all $k \geq 0$, the k^{th} intrinsic volume of a convex body is independent of the ambient dimension [20]. Thus, for $0 \leq k \leq a$, the k^{th} intrinsic volume of $\hat{\mathcal{S}}_{a+\gamma}$ is exactly the same as that of $\mathcal{S}_a$. For $a+1 \leq k \leq a+\gamma$, the k^{th} intrinsic volume of $\hat{\mathcal{S}}_{a+\gamma}$ equals 0. The sequence of intrinsic volumes of $\hat{S}_{a+\gamma}$ may therefore be considered to be simply μ_a. In addition, note that for all $m \geq 1$,

$$\underbrace{\hat{\mathcal{S}}_{a+\gamma} \times \cdots \times \hat{\mathcal{S}}_{a+\gamma}}_{m} \subseteq \mathcal{S}_{m(a+\gamma)},$$

which implies

$$\underbrace{\mu_a \star \cdots \star \mu_a}_{m} \leq \mu_{m(a+\gamma)}.$$

This leads us to define the new sequence $\hat{\mu}_n$ as

$$\hat{\mu}_n = \underbrace{\mu_a \star \cdots \star \mu_a}_{\lfloor \frac{n}{a+\gamma} \rfloor} := \mu_a^{\star \lfloor \frac{n}{a+\gamma} \rfloor}.$$

Clearly $\hat{\mu}_n \leq \mu_n$. Define $\hat{G}_n(t)$ as follows,

$$\hat{G}_n(t) = \log \sum_{j=0}^{n} \hat{\mu}_n(j) e^{jt},$$

and consider the limit

$$\lim_{n\to\infty} \frac{1}{n}\hat{G}_n(t) = \lim_{n\to\infty} \frac{1}{n} \lfloor \frac{n}{a+\gamma} \rfloor G_a(t) \tag{C.74}$$

$$= \frac{G_a(t)}{a+\gamma}. \tag{C.75}$$

Applying the Gärtner-Ellis theorem, stated in Theorem 2.0.1, for $\{\hat{\mu}_n\}$ and noting that $\frac{G_a(t)}{a+\gamma}$ is differentiable, we get the lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{\mu}_{n/n}(F) \geq -\inf_{x\in F} \left(\frac{G_a(t)}{a+\gamma}\right)^* (x),$$

which implies

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(F) \geq -\inf_{x\in F} \frac{a}{a+\gamma} g_a^* \left(\frac{a+\gamma}{a} x\right).$$

We claim that $\inf_{x\in F} \frac{a}{a+\gamma} g_a^* \left(\frac{a+\gamma}{a} x\right)$ converges to $\inf_{x\in F} \Lambda^*(x)$. Let $\epsilon > 0$. We can rewrite the infimum as

$$\inf_{x\in F} \frac{a}{a+\gamma} g_a^* \left(\frac{a+\gamma}{a} x\right) = \inf_{y\in \frac{a+\gamma}{a} F} \frac{a}{a+\gamma} g_a^*(y).$$

Using Theorem 2.1.4, we know that $\{g_n^*\}$ converges uniformly Λ^* over $[0,1]$. By the converse of the Arzela-Ascoli theorem, we have that g_n^* are uniformly bounded and equicontinuous. Let $\delta > 0$ be such that

$$|\Lambda^*(x) - \Lambda^*(y)| < \epsilon/3 \text{ whenever } |x-y| < \delta. \tag{C.76}$$

Let M be a uniform bound on $|g_n^*(\cdot)|$. Choose A_0 such that for all $a > A_0$,

$$\frac{\gamma}{a+\gamma} M < \epsilon/3. \tag{C.77}$$

Choose A_1 such that for all $a > A_1$,

$$||g_a^* - \Lambda^*||_\infty < \epsilon/3. \tag{C.78}$$

Choose A_2 such that for all $a > A_2$,

$$\frac{\gamma}{a+\gamma} < \delta. \tag{C.79}$$

Choose A_3 such that for all $a > A_3$,

$$\frac{a+\gamma}{a} F \cap [0,1] \neq \phi. \tag{C.80}$$

Now for all $a > \max(A_0, A_1, A_2, A_3)$,

$$\begin{aligned}
\left| \inf_{y \in \frac{a+\gamma}{a} F} \frac{a}{a+\gamma} g_a^*(y) - \inf_{y \in F} \Lambda^*(y) \right| &\leq \left| \inf_{y \in \frac{a+\gamma}{a} F} \frac{a}{a+\gamma} g_a^*(y) - \inf_{y \in \frac{a+\gamma}{a} F} g_a^*(y) \right| \\
&\quad + \left| \inf_{y \in \frac{a+\gamma}{a} F} g_a^*(y) - \inf_{y \in \frac{a+\gamma}{a} F} \Lambda^*(y) \right| \\
&\quad + \left| \inf_{y \in \frac{a+\gamma}{a} F} \Lambda^*(y) - \inf_{y \in F} \Lambda^*(y) \right| \\
&\overset{(a)}{<} \epsilon/3 + \epsilon/3 + \epsilon/3 \\
&= \epsilon.
\end{aligned}$$

By the relation (C.80), all the infimums involved in the above sequence of inequalities are finite. In step (a), the first term is less that $\epsilon/3$ by inequality (C.77), the second term is less that $\epsilon/3$ by inequality (C.78), and the last term is less that $\epsilon/3$ by inequality (C.79). This completes the proof of part 2 of Lemma 3.8.5, and thus completes the proof of Lemma 3.8.5.

C.4.5 Proof of Lemma 3.8.6

Note that the claims in points 1 and 2 immediately imply 3, since $f_n^\nu = a_n + b_n^\nu$.

We'll first prove the claim in point 2. We start by proving pointwise convergence of $\{b_n^\nu(\cdot)\}$. Recall the expression for $b_n^\nu(\theta)$,

$$b_n^\nu(\theta) = \frac{1}{n} \log \frac{\pi^{n\theta/2}}{\Gamma(n\theta/2+1)} (n\nu)^{n\theta/2}.$$

For $\theta = 0$, this convergence is obvious. Let $\theta > 0$. We use the approximation

$$\log \Gamma(z) = z \log z - z + O(\log z),$$

and get that

$$b_n^\nu(n\theta) = \frac{1}{n}\left[\frac{n\theta}{2}\log \pi n\nu - \frac{n\theta}{2}\log\frac{n\theta}{2e} + O(\log n\theta)\right] \tag{C.81}$$

$$= \frac{1}{n}\left[\frac{n\theta}{2}\log\frac{2\pi e\nu}{\theta} + O(\log n\theta)\right] \tag{C.82}$$

$$= \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta} + \frac{O(\log n\theta)}{n}. \tag{C.83}$$

Taking the limit as $n \to \infty$, the pointwise convergence of b_n^ν follows. Concavity of b_n^ν from point 2 of Lemma 3.8.4, combined with Lemma A.1.1 then implies uniform convergence.

We shall now prove point 1. We start by showing the pointwise convergence of $a_n(\theta)$ to $-\Lambda^*(1-\theta)$, or equivalently the convergence of $a_n(1-\theta)$ to $-\Lambda^*(\theta)$. Note that convergence at the boundary points is already known. Let $\theta_0 \in (0,1)$. For ease of notation, we denote

$$\chi(\theta) := -\Lambda^*(\theta)$$
$$\bar{a}_n(\theta) := a_n(1-\theta).$$

Note that $\bar{a}_n$ is linearly interpolated from its values at j/n, where $\bar{a}_n(j/n) = \frac{1}{n}\log\mu_n(j)$. Let $\epsilon > 0$ be given. The function χ, being continuous on the bounded interval $[0,1]$, is uniformly continuous. Choose $\delta > 0$ such that

$$|\chi(x) - \chi(y)| < \epsilon, \quad \text{whenever} \quad |x-y| < \delta.$$

Choose $N_0 > 1/(\delta/3)$, and divide the interval $[0,1]$ into the the N_0 intervals $I_j := \left[\frac{j}{N_0}, \frac{j+1}{N_0}\right]$ for $0 \le j \le N_0 - 1$. Note that each interval has length less than $\delta/3$. Without loss of generality, let θ_0 lie in the interior of the k-th interval (we can always choose a different value of N_0 to make sure θ_0 does not lie on the boundary of any interval). Thus,

$$\frac{k-1}{N_0} < \theta_0 < \frac{k}{N_0}.$$

Lemma 3.8.5 along with the continuity of χ imply that

$$\lim_{n\to\infty}\frac{1}{n}\log\mu_n(I_j) = \sup_{\theta\in I_j}\chi(\theta). \tag{C.84}$$

For $n > 2/\min\left(\theta_0 - \frac{k-1}{N_0}, \frac{k}{N_0} - \theta_0\right)$, there exists an i such that

$$\frac{k-1}{N_0} < \frac{i}{n} < \theta_0 < \frac{i+1}{n} < \frac{k}{N_0}. \tag{C.85}$$

Thus for some $\lambda > 0$, we can write

$$\bar{a}_n(\theta_0) = \lambda \frac{1}{n} \log \mu_{n/n}(i/n) + (1-\lambda)\frac{1}{n} \log \mu_{n/n}((i+1)/n), \tag{C.86}$$

and obtain the inequality

$$\begin{aligned}
\bar{a}_n(\theta_0) &= \lambda \frac{1}{n} \log \mu_{n/n}(i/n) + (1-\lambda)\frac{1}{n} \log \mu_{n/n}((i+1)/n) && \text{(C.87)} \\
&\leq \max\left(\frac{1}{n} \log \mu_{n/n}(i/n), \frac{1}{n} \log \mu_{n/n}((i+1)/n)\right) && \text{(C.88)} \\
&\leq \frac{1}{n} \log \mu_{n/n}(I_k). && \text{(C.89)}
\end{aligned}$$

Thus we have the upper bound

$$\begin{aligned}
\limsup_n \bar{a}_n(\theta_0) &\leq \lim_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(I_k) && \text{(C.90)} \\
&= \sup_{\theta \in I_k} \chi(\theta) && \text{(C.91)} \\
&\overset{(a)}{\leq} \chi(\theta_0) + \epsilon && \text{(C.92)}
\end{aligned}$$

where (a) follows from the choice of N_0 and uniform continuity of χ.

Define

$$\hat{\theta}_n(j) = \arg \sup_{\frac{i}{n} \text{ s.t. } \frac{i}{n} \in I_j} \mu_{n/n}\left(\frac{i}{n}\right).$$

As

$$\mu_{n/n}(\hat{\theta}_n(j)) \leq \mu_{n/n}(I_j) \leq \left(\frac{n}{N_0} + 2\right) \mu_{n/n}(\hat{\theta}_n(j)) \leq n\mu_{n/n}(\hat{\theta}_n(j)),$$

it is easy to see that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(\hat{\theta}_n(j)) &= \lim_{n\to\infty} \frac{1}{n} \log \mu_{n/n}(I_j) && \text{(C.93)} \\
&= \sup_{\theta \in I_j} \chi(\theta). && \text{(C.94)}
\end{aligned}$$

Note that

$$\sup_{\theta \in I_j} \bar{a}_n(\theta) \geq \frac{1}{n} \log \mu_{n/n}(\hat{\theta}_n(j)).$$

This implies that for the intervals I_{k-1} and I_{k+1},

$$\liminf_{n\to\infty} \left[\sup_{\theta \in I_{k-1}} \bar{a}_n(\theta)\right] \geq \sup_{\theta \in I_{k-1}} \chi(\theta) \geq \chi(\theta_0) - \epsilon \tag{C.95}$$

$$\liminf_{n\to\infty} \left[\sup_{\theta \in I_{k+1}} \bar{a}_n(\theta)\right] \geq \sup_{\theta \in I_{k+1}} \chi(\theta) \geq \chi(\theta_0) - \epsilon. \tag{C.96}$$

Since $\bar{a}_n(\theta)$ is concave, this implies

$$\bar{a}_n(\theta_0) \geq \min(\sup_{\theta \in I_{k-1}} \bar{a}_n(\theta), \sup_{\theta \in I_{k+1}} \bar{a}_n(\theta)). \tag{C.97}$$

Taking the lim inf on both sides,

$$\liminf_{n\to\infty} \bar{a}_n(\theta_0) \geq \chi(\theta_0) - \epsilon. \tag{C.98}$$

Inequalities (C.90) and (C.98) prove the pointwise convergence of $\bar{a}_n(\theta_0)$ to $\chi(\theta_0)$. Concavity of a_n from point 2 of Lemma 3.8.4, combined with Lemma A.1.1 then implies uniform convergence.

C.4.6 Proof of Lemma 3.8.7

By Lemma 3.8.6, the sequence of functions $\{f_n^\nu\}$ converges to f^ν uniformly. Using the converse of the Arzela-Ascoli theorem, this implies that the family of functions $\{f_n^\nu\}$ is equicontinuous. Let $\epsilon > 0$ be given. Choose N large such that $|f_n^\nu(x) - f_n^\nu(y)| < \epsilon/2$ if $|x-y| < 1/N$. This implies that for all $n > N$,

$$\max_\theta f_n^\nu(\theta) \geq f_n^\nu(\hat{\theta}_n) > \max_\theta f_n^\nu(\theta) - \epsilon/2. \tag{C.99}$$

Using the uniform convergence of $\{f_n^\nu\}$, we choose M large enough such that $||f^\nu - f_n^\nu||_\infty < \epsilon/2$ for all $n > M$. Let $L = \max(M, N)$. For all $n > L$, we have

$$\max_\theta f^\nu(\theta) + \epsilon/2 > \max_\theta f_n^\nu(\theta) \geq f_n^\nu(\hat{\theta}_n) \geq \max_\theta f_n^\nu(\theta) - \epsilon/2 \geq \max_\theta f^\nu(\theta) - \epsilon,$$

and thus

$$|f_n^\nu(\hat{\theta}_n) - \max_\theta f^\nu(\theta)| < \epsilon.$$

This concludes the proof.

C.4.7 Proof of Lemma 3.8.8

Recall that

$$\mathrm{Vol}(\mathcal{S}_n(\sigma,\rho)^n \oplus B_n(\sqrt{n\nu})) = \sum_{j=0}^{n} e^{n f_n^\nu(j/n)}.$$

We have the trivial bounds

$$e^{n f_n^\nu(\hat{\theta}_n)} \leq \mathrm{Vol}(\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu})) \leq (n+1) e^{n f_n^\nu(\hat{\theta}_n)} \tag{C.100}$$

implying

$$f_n^\nu(\hat{\theta}_n) \leq \frac{1}{n} \log \mathrm{Vol}(\mathcal{S}_n(\sigma,\rho) \oplus B_n(\sqrt{n\nu})) \leq \frac{\log(n+1)}{n} + f_n^\nu(\hat{\theta}_n). \tag{C.101}$$

Taking the limit in n, we obtain

$$\lim_{n\to\infty} f_n^\nu(\hat{\theta}_n) = \ell(\nu). \tag{C.102}$$

An application of Lemma 3.8.7 gives

$$\ell(\nu) = \sup_\theta f^\nu(\theta). \tag{C.103}$$

C.4.8 Proof of Lemma 3.8.9

Recall the expression of $f^\nu(\theta)$:

$$f^\nu(\theta) = -\Lambda^*(1-\theta) + \frac{\theta}{2}\log\frac{2\pi e\nu}{\theta}. \tag{C.104}$$

Suppose $\limsup_{\nu\to 0}\theta^*(\nu) = \eta > 0$. Choose a sequence $\{\nu_n\}$ such that

$$\begin{aligned} &\lim_{n\to\infty} \nu_n = 0 && \text{(C.105)}\\ &\theta^*(\nu_n) > \frac{\eta}{2} \text{ for all } n \geq 1. && \text{(C.106)}\end{aligned}$$

We have that for all $\nu > 0$,

$$\ell(\nu) = \sup_\theta f^\nu(\theta) \geq f^\nu(0) = -\Lambda^*(1) = v(\sigma,\rho).$$

Thus,

$$\begin{aligned} v(\sigma,\rho) &\leq \ell(\nu_n) && \text{(C.107)}\\ &= f^{\nu_n}(\theta^*(\nu_n)) && \text{(C.108)}\\ &= -\Lambda^*(\theta^*(\nu_n)) + \frac{\theta^*(\nu_n)}{2}\log\frac{2\pi e\nu_n}{\theta^*(\nu_n)} && \text{(C.109)}\\ &\leq \sup_\theta\left[-\Lambda^*(1-\theta) + \frac{\theta}{2}\log\frac{2\pi e}{\theta}\right] + \frac{\theta^*(\nu_n)}{2}\log\nu_n && \text{(C.110)}\\ &\overset{(a)}{\leq} C + \frac{\theta^*(\nu_n)}{2}\log\nu_n && \text{(C.111)}\\ &\overset{(b)}{\leq} C + \frac{\eta}{4}\log\nu_n && \text{(C.112)}\end{aligned}$$

where in (a), C is a constant and in (b) we assume $\log\nu_n < 0$. Taking the limit as $n\to\infty$, we get that

$$v(\sigma,\rho) \leq \lim_{n\to\infty} C + \frac{\eta}{4}\log\nu_n = -\infty, \tag{C.113}$$

which is a contradiction. Thus, it must be that $\limsup_{\nu\to 0}\theta^*(\nu) = 0$.

C.5 Proof for Section 3.9

C.5.1 Proof of Lemma 3.9.4

We shall first show that as $\theta \to 1$,

$$\Lambda^*(d) - \Lambda^*(d\theta) + d(1-\theta)\log(1-\theta) = O(1-\theta). \tag{C.114}$$

Recall that

$$\Lambda(t) = \log\left(\sum_{j=0}^{d} \alpha_j e^{jt}\right),$$

and $\Lambda^*(d\theta)$ is given by

$$\Lambda^*(d\theta) = \sup_t d\theta t - \log\left(\sum_{j=0}^{d} \alpha_j e^{jt}\right). \tag{C.115}$$

Let $t^*(\theta)$ be

$$t^*(\theta) = \arg\sup_t d\theta t - \log\left(\sum_{j=0}^{d} \alpha_j e^{jt}\right). \tag{C.116}$$

We shall sometimes refer to $t^*(\theta)$ simply by t^* when the argument is understood. We have that t^* satisfies

$$d\theta = \frac{\sum_{j=0}^{d} j\alpha_j e^{jt^*}}{\sum_{j=0}^{d} \alpha_j e^{jt^*}}, \tag{C.117}$$

which implies

$$d(1-\theta) = \frac{\sum_{j=0}^{d} (d-j)\alpha_j e^{jt^*}}{\sum_{j=0}^{d} \alpha_j e^{jt^*}}. \tag{C.118}$$

Choose a θ such that $d\theta > \frac{d}{dt}\Lambda(t)\Big|_{t=0}$ to ensure that $t^*(\theta) > 0$. We have the inequality,

$$\frac{\alpha_{d-1}e^{(d-1)t^*}}{M_1 e^{dt^*}} \le \frac{\sum_{j=0}^{d} (d-j)\alpha_j e^{jt^*}}{\sum_{j=0}^{d} \alpha_j e^{jt^*}} \le \frac{M_2 e^{(d-1)t^*}}{\alpha_d e^{nt^*}}, \tag{C.119}$$

which implies

$$\frac{\alpha_{d-1}}{dM_1} e^{-t^*} \le (1-\theta) \le \frac{M_2}{d\alpha_d} e^{-t^*}. \tag{C.120}$$

where $M_1 = (d+1)\max_j \alpha_j$ and $M_2 = (d+1)\max_j (d-j)\alpha_j$. Taking logarithms, we get

$$c_1 - t^* \le \log(1-\theta) \le c_2 - t^*, \tag{C.121}$$

for constants c_1 and c_2. Now let us consider the difference $\Lambda^*(d) - \Lambda^*(d\theta)$. Since $\Lambda^*(d) = -\log \alpha_d$, we can write $\Lambda^*(d) - \Lambda^*(d\theta)$ as

$$\Lambda^*(d) - \Lambda^*(d\theta) = -\log \alpha_d - \sup_t d\theta t - \Lambda(t) \tag{C.122}$$

$$= -\log \alpha_d - d\theta t^* + \Lambda(t^*) \tag{C.123}$$

$$= dt^*(1-\theta) - \log \alpha_d e^{dt^*} + \log \sum_j \alpha_j e^{jt^*} \tag{C.124}$$

$$= dt^*(1-\theta) + \log\left(1 + \frac{\sum_{j=0}^{d-1} \alpha_j e^{jt^*}}{\alpha_d e^{dt^*}}\right). \tag{C.125}$$

Adding $d(1-\theta)\log(1-\theta)$ to both sides, we get

$$\Lambda^*(d) - \Lambda^*(d\theta) + d(1-\theta)\log(1-\theta) = \underbrace{d(1-\theta)[t^* + \log(1-\theta)]}_{O(1-\theta) \text{ by (C.121)}} + \underbrace{\log\left(1 + \frac{\sum_{j=0}^{d-1} \alpha_j e^{jt^*}}{\alpha_d e^{dt^*}}\right)}_{O(e^{-t^*}), \text{ which equals } O(1-\theta) \text{ by (C.120)}}$$

$$= O(1-\theta) \tag{C.126}$$

We shall now study the asymptotics of $1 - \theta^*(\nu)$ as $\nu \to 0$. We have that $\theta^*(\nu)$ satisfies

$$\frac{d}{d\theta}\left[-\Lambda^*(d\theta) + \frac{d(1-\theta)}{2}\log\frac{2\pi e\nu}{1-\theta}\right]\Bigg|_{\theta=\theta^*} = 0. \tag{C.127}$$

This gives

$$-\Lambda^{*'}(d\theta^*) - \frac{1}{2}\log 2\pi e\nu + \frac{1}{2} + \frac{1}{2}\log(1-\theta^*) = 0, \tag{C.128}$$

which implies

$$-\Lambda^{*'}(d\theta^*) + \frac{1}{2}\log(1-\theta^*) = \frac{1}{2}\log 2\pi\nu. \tag{C.129}$$

Now $\Lambda^{*'}(d\theta^*)$ is simply $t^*(\theta^*)$ as defined in equation (C.116). By inequality (C.120), there exist some constants c_1 and c_2 such that

$$-\log(1-\theta^*) + c_1 \le t^*(\theta^*) \le -\log(1-\theta^*) + c_2. \tag{C.130}$$

Substituting in equation (C.129), there exist some constants c_1 and c_2 such that

$$3\log(1-\theta^*) + c_1 \leq \log\nu \leq 3\log(1-\theta^*) + c_2, \tag{C.131}$$

which implies there exist some constants c_1 and c_2 such that

$$c_1 \leq \frac{1-\theta^*}{\nu^{1/3}} \leq c_2. \tag{C.132}$$

Thus, $1-\theta^*$ is $O(\nu^{1/3})$. This combined with equation (C.126) completes the proof of Lemma 3.9.4.

C.5.2 Proof of Lemma 3.9.5

From the inequality (C.132), we have that there exist some constants c_1 and c_2

$$c_1 \leq \log\frac{2\pi e\nu}{(1-\theta^*)^3} \leq c_2. \tag{C.133}$$

We also have $1-\theta^*$ is $O(\nu^{1/3})$. Thus, it follows that $\frac{d(1-\theta^*)}{2}\log\frac{2\pi e\nu}{(1-\theta^*)^3} = O(\nu^{1/3})$.

C.5.3 Proof of Lemma 3.9.9

Since $\mathcal{A}^\nu(\theta)$ is a continuous function of θ on the compact set $[0,1]$, it is uniformly continuous. Choose $\delta > 0$ such that

$$|\mathcal{A}^\nu(x) - \mathcal{A}^\nu(y)| < \frac{\eta}{10}, \quad \text{if} \quad |x-y| < \delta.$$

Divide the interval $[0,1]$ into M equals parts $[\alpha_i, \alpha_{i+1}]$ for $0 \leq i \leq M-1$, with $\alpha_0 = 0$ and $\alpha_M = 1$. We choose M such that each interval has length less than δ. Since $\{f_n^\nu\}$ converge pointwise on $[\alpha_1, \alpha_{M-1}]$ to $\mathcal{A}(\nu,\theta)$, we can use Lemma A.1.1 to conclude that this convergence is uniform on $[\alpha_1, \alpha_{M-1}]$. Choose N such that for all $\theta \in [\alpha_1, \alpha_{M-1}]$ and for all $n > N$, we have

$$|\mathcal{A}^\nu(\theta) - f_n^\nu(\theta)| < \frac{\eta}{10}. \tag{C.134}$$

For $\theta \in [0, \alpha_1]$, by uniform continuity of $\mathcal{A}^\nu(\theta)$ we have

$$\mathcal{A}^\nu(\alpha_1) - \frac{n}{10} \leq \mathcal{A}^\nu(\theta) < \mathcal{A}^\nu(\alpha_1) + \frac{n}{10} \tag{C.135}$$

For $\theta \in [0, \alpha_1]$, we can use the concavity of f_n^ν to upper bound the value of $f_n^\nu(\theta)$ as follows. We write α_1 as a linear combination of θ and α_2, and use the concavity of f_n^ν

to get

$$
\begin{aligned}
f_n^\nu(\alpha_1) &\geq \frac{\alpha_2-\alpha_1}{\alpha_2-\theta} f_n^\nu(\theta) + \frac{\alpha_1-\theta}{\alpha_2-\theta} f_n^\nu(\alpha_2) \ , \\
\implies \frac{\alpha_2-\theta}{\alpha_2-\alpha_1} f_n^\nu(\alpha_1) &- \frac{\alpha_1-\theta}{\alpha_2-\alpha_1} f_n^\nu(\alpha_2) \geq f_n^\nu(\theta) \ , \\
\implies \sup_{\theta\in(\alpha_0,\alpha_1)} \frac{\alpha_2-\theta}{\alpha_2-\alpha_1} f_n^\nu(\alpha_1) &- \frac{\alpha_1-\theta}{\alpha_2-\alpha_1} f_n^\nu(\alpha_2) \geq f_n^\nu(\theta) \ .
\end{aligned}
$$

Note that since the LHS is linear in x, the supremum occurs at one of the endpoints of the interval.

$$
\begin{aligned}
f_n^\nu(\theta) &\leq \max\left(f_n^\nu(\alpha_1), 2f_n^\nu(\alpha_1) - f_n^\nu(\alpha_2)\right) \ , \\
&\leq \max(\mathcal{A}^\nu(\alpha_1)+\eta/10, 2(\mathcal{A}^\nu(\alpha_1)+\eta/10) - (\mathcal{A}^\nu(\alpha_2)-\eta/10) \ ) \\
&\leq \max(\mathcal{A}^\nu(\alpha_1)+\eta/10, 2\mathcal{A}^\nu(\alpha_1) - \mathcal{A}^\nu(\alpha_2)+3\eta/10) \\
&\leq \max(\mathcal{A}^\nu(\alpha_1)+\eta/10, 2\mathcal{A}^\nu(\alpha_1) - \mathcal{A}^\nu(\alpha_1)+\eta/10+3\eta/10) \ , \\
&= \mathcal{A}^\nu(\alpha_1)+4\eta/10 \ , \\
&\leq \mathcal{A}^\nu(\theta)+5\eta/10 \ , \\
&< \mathcal{A}^\nu(\theta)+\eta.
\end{aligned} \tag{C.136}
$$

We can use a similar strategy for $\theta \in [\alpha_{M-1}, \alpha_M]$ to bound $f^\nu(\theta)$ from above by $\mathcal{A}^\nu(\theta)+\eta$. Thus we conclude that for all $\theta \in [0,1]$,

$$
f_n^\nu(\theta) < \mathcal{A}^\nu(\theta)+\eta. \tag{C.137}
$$

Appendix D

Proofs for Chapter 4

D.1 Proofs for Section 4.1

D.1.1 Proof of Lemma 4.1.2

Note that $\mu_m^\epsilon \star \mu_n^\epsilon(m+n) = \mu_n^\epsilon(n)\mu_m^\epsilon(m)$ and $\mu_m^\epsilon \star \mu_n^\epsilon(0) = \mu_n^\epsilon(0)\mu_m^\epsilon(0)$. Thus the existence of the limits defining α and β is given by sub-additivity. Existence of limit defining γ follows from the equality $\gamma = \Lambda^\epsilon(0)$.

Proof of $\alpha < \infty$: Note that $\mu_n(n)$ is simply the volume of the typical set $\mathcal{T}_n^\epsilon$. Since $P(\mathcal{T}_n^\epsilon) \le 1$, we have

$$|\mathcal{T}_n^\epsilon| \le e^{n(h(X)+\epsilon)}, \tag{D.1}$$

so $\alpha \le h(X) + \epsilon$, and is therefore finite.

Proof of $\beta < \infty$ and $\mu_n^\epsilon(0), \mu_n^\epsilon(n) > 0$: The value of $\mu_n(0)$ is the Euler characteristic, which equals 1 when $\mathcal{T}_n^\epsilon$ is non-empty. We show that for every $n \ge 1$, the set $\mathcal{T}_n^\epsilon$ has a nonempty interior; i.e., $\mathrm{Vol}(\mathcal{T}_n^\epsilon) = \mu_n(n) > 0$. Let $M = \max_x p_X(x) = e^{-\min_x \Phi(x)}$. Note that the set of minimizers of Φ is a nonempty set, since $\Phi \to +\infty$ as $|x| \to +\infty$. Let x^* be any such minimizer of Ψ. For the point $(x^*, \dots, x^*) \in \mathbb{R}^n$, we have

$$\sum_{i=1}^n \Phi(x_i) = -n \log M.$$

We also have the inequality

$$-h(X) = \int_{\mathbb{R}} p_X(x) \log p_X(x) dx \le \int_{\mathbb{R}} p_X(x) \log M dx = \log M.$$

Thus, for the point $(x^*, \dots, x^*)$, we have

$$\sum_{i=1}^n \Phi(x_i) = -n \log M < n(h(X) + \epsilon),$$

so $(x^*, \ldots, x^*) \in \mathcal{T}_n^\epsilon$. By the continuity of Φ at x^*, we conclude that $\mathcal{T}_n^\epsilon$ has a nonempty interior.

Proof of $\gamma < \infty$: Since $\Phi \to \pm\infty$ as $|x| \to \pm\infty$, we may find constants $c_1 > 0$ and c_2 such that

$$\Phi(x) \geq c_1|x| + c_2, \quad \text{for all} \quad x \in \mathbb{R}. \tag{D.2}$$

We start by showing that for $A = \frac{h(X)+\epsilon-c_2}{c_1}$, the sequence of regular crosspolytopes $\{C_n\}_{n=1}^\infty$ defined by

$$C_n := \{x^n \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i| \leq An\}$$

satisfies the containment

$$\mathcal{T}_n^\epsilon \subseteq C_n, \quad \text{for} \quad n \geq 1. \tag{D.3}$$

For $x^n \in \mathcal{T}_n^\epsilon$, using definition (4.4) and inequality (D.2), we have

$$\sum_{i=1}^n (c_1|x_i| + c_2) \leq \sum_{i=1}^n \Phi(x_i) \leq n(h(X) + \epsilon),$$

implying that

$$\sum_{i=1}^n |x_i| \leq n\left(\frac{h(X) + \epsilon - c_2}{c_1}\right),$$

so $x^n \in C_n$. Hence, $\mathcal{T}_n^\epsilon \subseteq C_n$, as claimed. Let the intrinsic volumes of C_n be $\hat{\mu}_n(\cdot)$. Note that $\mu_n(i) \leq \hat{\mu}_n(i)$ for all $0 \leq i \leq n$, by the containment (D.3). Thus, $\gamma \leq \hat{\gamma}$, where

$$\hat{\gamma} := \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{i=0}^n \hat{\mu}_n(i)\right). \tag{D.4}$$

We claim that $\hat{\gamma} < \infty$. Define

$$\hat{G}_n(t) = \log \sum_{i=0}^n \hat{\mu}_n(i) e^{it}, \quad \text{and} \quad \hat{g}_n(t) = \frac{\hat{G}_n(t)}{n}.$$

Note that the sequence $\{C_n\}$ is super-convolutive, so $\hat{g}_n(t)$ converges pointwise. In particular, for $t = 0$, we have

$$\hat{\gamma} = \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{i=0}^n \mu_i(C_n)\right) \quad \text{exists, and is possibly} \quad +\infty.$$

The i-th intrinsic volume of C_n is given by [2]

$$\hat{\mu}_n(i) = \begin{cases} 2^{i+1}\binom{n}{i+1}\frac{\sqrt{i+1}}{i!}\frac{(nA)^i}{\sqrt{\pi}}\times \\ \qquad \int_0^\infty e^{-x^2}\left(\frac{2}{\sqrt{\pi}}\int_0^{x/\sqrt{i+1}} e^{-y^2}dy\right)^{n-i-1} dx \\ \qquad\qquad\qquad\qquad\qquad \text{if } \; i \le n-1 \\ \frac{2^n}{n!}(nA)^n \qquad\qquad\qquad \text{if } \; i = n. \end{cases}$$

Note that

$$\begin{aligned} &\int_0^\infty e^{-x^2}\left(\frac{2}{\sqrt{\pi}}\int_0^{x/\sqrt{i+1}} e^{-y^2}dy\right)^{n-i-1} dx \\ &\le \int_0^\infty e^{-x^2}\left(\frac{2}{\sqrt{\pi}}\int_0^{\infty} e^{-y^2}dy\right)^{n-i-1} dx \\ &\le \int_0^\infty e^{-x^2}dx \\ &= \frac{\sqrt{\pi}}{2}. \end{aligned}$$

Thus, for $0 \le i \le n-1$,

$$\hat{\mu}_n(i) \le 2^{i+1}\binom{n}{i+1}\frac{\sqrt{i+1}}{i!}\frac{(nA)^i}{\sqrt{\pi}} \times \frac{\sqrt{\pi}}{2} \tag{D.5}$$

$$= 2^{i}\binom{n}{i+1}\frac{\sqrt{i+1}}{i!}(nA)^i \tag{D.6}$$

$$\le 2^n \times 2^n \times \sqrt{n+1} \times A^i \times \frac{n^i}{i!} \tag{D.7}$$

$$\le 2^{2n}\sqrt{n+1} \times \max(1, A^n) \times \frac{n^n}{n!}. \tag{D.8}$$

We may check that the inequality also holds for $i = n$. Hence,

$$\begin{aligned} &\frac{1}{n}\log\left(\sum_{i=0}^{n}\mu_i(C_n)\right) \\ &\le \frac{1}{n}\log\left((n+1)\times 2^{2n}\sqrt{n+1}\times\max(1,A^n)\times\frac{n^n}{n!}\right). \end{aligned}$$

Taking the limit as $n \to \infty$, we obtain

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\log\left(\sum_{i=0}^{n}\mu_i(C_n)\right) &\le 2\log 2 + \max(0, \log A) + 1 \\ &< \infty. \end{aligned}$$

This shows that $\hat{\gamma}$ is finite, and therefore γ is finite.

D.2 Proofs for Section 4.2

D.2.1 Proof of Lemma 4.2.2

Without loss of generality, take $a = 0$ and $b = 1$. Since f is the pointwise limit of concave functions, it is also concave. The continuity of f is not obvious a priori: it could be discontinuous at the endpoints 0 and 1. Let $f(0) = \ell_0$ and $f(1) = \ell_1$. For any $n \geq 1$, the function f_n is lower-bounded by the line joining $(0, \ell_0)$ and $(1, \ell_1)$. Call this lower bound $L(\theta)$, for $\theta \in [0, 1]$. We prove continuity at 0, by showing that for $\eta > 0$, there exists a $\delta > 0$ such that for $\theta \in [0, \delta)$, we have $|f(\theta) - \ell_0| < \eta$. Pick N large enough such that

$$f_N(0) - \ell_0 < \eta/2.$$

The function f_N is continuous on $[0, 1]$, so there exists $\delta_1 > 0$ such that for $\theta \in [0, \delta_1)$, we have $|f_n(\theta) - f_n(0)| < \eta/2$. Now pick a δ_2 such that $|L(\theta) - \ell_0| < \eta/2$, for $\theta \in [0, \delta_2)$. Let $\delta = \min(\delta_1, \delta_2)$. For $n > N$, we have

$$L(\theta) \leq f_n(\theta) \leq f_N(\theta).$$

Thus, for $\theta \in [0, \delta)$, we obtain

$$f_n(\theta) \leq f_N(\theta) \leq f_N(0) + \eta/2 \leq \ell_0 + \eta,$$

and

$$f_n(\theta) \geq L(\theta) \geq \ell_0 - \eta/2.$$

Thus, for all $n > N$ and $\theta \in [0, \delta)$, we have

$$\ell_0 - \eta/2 \leq f_n(\theta) \leq \ell_0 + \eta.$$

Taking the limit as $n \to \infty$, we conclude that for $\theta \in [0, \delta)$,

$$\ell_0 - \eta/2 \leq f(\theta) \leq \ell_0 + \eta,$$

implying continuity at 0. Continuity at 1 follows similarly.

D.2.2 Proof of Lemma 4.2.4

From inequality (D.7), for all i we have

$$\hat{\mu}_n(i) \leq 2^i \binom{n}{i+1} \frac{\sqrt{i+1}}{i!} (nA)^i \tag{D.9}$$

Substituting $i = \lfloor n\theta \rfloor$, taking $1/n$ times the logarithms on both sides and taking the limit as $n \to \infty$, we obtain

$$-\chi^*(\theta) \leq \theta \log 2A + H(\theta) - \theta \log \frac{\theta}{e}. \tag{D.10}$$

By concavity of $-\chi^*$, and since $-\chi^*(0) = 0$, we know that

$$\lim_{\theta \to 0} -\chi^*(\theta) \geq 0. \tag{D.11}$$

Taking the limit as $\theta \to 0$ in equation (D.10), we obtain the lower bound

$$\lim_{\theta \to 0} -\chi^*(\theta) \leq 0. \tag{D.12}$$

Thus, we must have

$$\lim_{\theta \to 0} -\chi^*(\theta) = 0, \tag{D.13}$$

and this shows that $-\chi^*$ is continuous at 0.

Bibliography

[1] Philip M Anselone and Joel Davis. *Collectively compact operator approximation theory and applications to integral equations*. Prentice-Hall Englewood Cliffs, NJ, 1971.

[2] Ulrich Betke and Martin Henk. “Intrinsic volumes and lattice points of crosspolytopes”. In: *Monatshefte für Mathematik* 115.1-2 (1993), pp. 27–33.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[4] Stefano Campi and Paolo Gronchi. “Estimates of Loomis–Whitney type for intrinsic volumes”. In: *Advances in Applied Mathematics* 47.3 (2011), pp. 545–561.

[5] Max Costa and Thomas Cover. “On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (Corresp.)” In: *Information Theory, IEEE Transactions on* 30.6 (1984), pp. 837–839.

[6] Max HM Costa. “A new entropy power inequality”. In: *IEEE Transactions on Information Theory* 31.6 (1985), pp. 751–760.

[7] T.M. Cover, J.A. Thomas, J. Wiley, et al. *Elements of information theory*. Vol. 6. Wiley Online Library, 1991.

[8] Rene L Cruz. “A calculus for network delay Part I: Network elements in isolation”. In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 114–131.

[9] Rene L Cruz. “A calculus of delay Part II: Network analysis”. In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 132–141.

[10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Vol. 2. Springer, 1998.

[11] RL Dobrushin. “General formulation of Shannons main theorem in information theory”. In: *Amer. Math. Soc. Trans* 33 (1963), pp. 323–438.

[12] Yishun Dong, Farzan Farnia, and Ayfer Özgür. “Near Optimal Energy Control and Approximate Capacity of Energy Harvesting Communication”. In: *arXiv preprint arXiv:1405.1156* (2014).

[13] Rick Durrett. *Probability: theory and examples*. Vol. 3. Cambridge University Press, 2010.

[14] R Gardner. "The Brunn-Minkowski inequality". In: *Bulletin of the American Mathematical Society* 39.3 (2002), pp. 355–405.

[15] John Hopcroft and Ravindran Kannan. "Foundations of Data Science". In: *Available online at http://research.microsoft.com/en-us/people/kannan/book-dec-30-2013.pdf* ().

[16] Varun Jog and Venkat Anantharam. "A Geometric Analysis of the AWGN channel with a (σ, ρ)-Power Constraint". In: *IEEE International Symposium on Information Theory (ISIT), 2015.*

[17] Varun Jog and Venkat Anantharam. "An energy harvesting AWGN channel with a finite battery". In: *IEEE International Symposium on Information Theory (ISIT), 2014*. IEEE. 2014, pp. 806–810.

[18] Varun Jog and Venkat Anantharam. "On the geometry of convex typical sets". In: *IEEE International Symposium on Information Theory (ISIT), 2015.*

[19] Daniel A Klain. "Invariant valuations on star-shaped sets". In: *advances in mathematics* 125.1 (1997), pp. 95–113.

[20] Daniel A Klain and Gian-Carlo Rota. *Introduction to geometric probability.* Cambridge University Press, 1997.

[21] Bo'az Klartag. "A central limit theorem for convex sets". In: *Inventiones mathematicae* 168.1 (2007), pp. 91–131.

[22] Wei Mao and Babak Hassibi. "On the capacity of a communication system with energy harvesting and a limited battery". In: *Proceedings of the 2013 International Symposium on Information Theory (ISIT)*. IEEE. 2013, pp. 1789–1793.

[23] Peter McMullen. "Inequalities between intrinsic volumes". In: *Monatshefte für Mathematik* 111.1 (1991), pp. 47–53.

[24] Joseph V Michalowicz, Jonathan M Nichols, and Frank Bucholtz. "Calculation of differential entropy for a mixed Gaussian distribution". In: *Entropy* 10.3 (2008), pp. 200–206.

[25] Omur Ozel and Sennur Ulukus. "Achieving AWGN Capacity Under Stochastic Energy Harvesting". In: *IEEE Transactions on Information Theory* 58.10 (2012), pp. 6471–6483.

[26] Grigoris Paouris, Peter Pivovarov, and Joel Zinn. "A central limit theorem for projections of the cube". In: *Probability Theory and Related Fields* 159.3-4 (2014), pp. 701–719.

[27] Carla Peri. "On relative isoperimetric inequalities". In: Aracne. 2001.

[28] Halsey Lawrence Royden and Patrick Fitzpatrick. *Real analysis, 4th edition.* Pearson, 2011.

[29] H.H. Schaefer and M.P.H. Wolff. *Topological Vector Spaces*. Graduate Texts in Mathematics. Springer New York, 1999. ISBN: 9780387987262. URL: http://books.google.com/books?id=9kXY742pABoC.

[30] Rolf Schneider. *Convex bodies: the Brunn-Minkowski theory*. Vol. 151. Cambridge University Press, 2013.

[31] Rolf Schneider and Wolfgang Weil. *Stochastic and integral geometry*. Springer, 2008.

[32] Shlomo Shamai and Israel Bar-David. "The capacity of average and peak-power-limited quadrature Gaussian channels". In: *IEEE Transactions on Information Theory* 41.4 (1995), pp. 1060–1071.

[33] C.E. Shannon. "A mathematical theory of communication, I and II". In: *Bell Syst. Tech. J* 27 (1948), pp. 379–423.

[34] Joel G Smith. "The information capacity of amplitude-and variance-constrained scalar gaussian channels". In: *Information and Control* 18.3 (1971), pp. 203–219.

[35] J Michael Steele. *Probability theory and combinatorial optimization*. Vol. 69. SIAM, 1997.

[36] Sujesha Sudevalayam and Purushottam Kulkarni. "Energy harvesting sensor nodes: Survey and implications". In: *Communications Surveys & Tutorials, IEEE* 13.3 (2011), pp. 443–461.

[37] S. Szarek and D. Voiculescu. "Shannon's entropy power inequality via restricted Minkowski sums". In: *Geometric aspects of functional analysis* (2000), pp. 257–262.

[38] Kaya Tutuncuoglu et al. "Binary energy harvesting channel with finite energy storage". In: *Proceedings of the 2013 International Symposium on Information Theory (ISIT)*. IEEE. 2013, pp. 1591–1595.

[39] Kaya Tutuncuoglu et al. "Improved capacity bounds for the binary energy harvesting channel". In: *Proceedings of the 2014 International Symposium on Information Theory (ISIT)*. IEEE. 2014, 976980.