# A STUDY OF INFORMATION CONTENT IN LINEAR AND METRIC SPACES

by

Dongbin Li

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics

Winter 2023

# A STUDY OF INFORMATION CONTENT IN LINEAR AND METRIC SPACES

by

Dongbin Li

Approved: ____________________________________________
Mark S. Gockenbach, Ph.D.
Chair of the Department of Mathematical Sciences

Approved: ____________________________________________
John A. Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: ____________________________________________
Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Mokshay M. Madiman, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Dominique Guillot, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Douglas Rizzolo, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Artem Zvavitch, Ph.D.
Member of dissertation committee

## ACKNOWLEDGEMENTS

After a long journey, I am about to receive my doctoral degree. I would like to use this opportunity to thank many people who helped me, encouraged me and accompanied me during my PhD studies.

I could not have undertaken this journey without the support of my advisor, Prof. Mokshay Madiman. I am extremely grateful to him for his invaluable patience, encouragement and guidance throughout my PhD studies.

I would like to express my deepest gratitude to my committee members, Prof. Dominique Guillot, Prof. Douglas Rizzolo and Prof. Artem Zvavitch, for carefully reading my thesis and providing many valuable suggestions.

The completion of my dissertation would not have been possible without the help of my collaborators. I am deeply indebted to my friend, my collaborator and my dear brother - Dr. Gautam Aishwarya. I would like to thank him for introducing me to our wonderful advisor Prof. Mokshay Madiman, for discussing math with me continuously, and also for offering so much help during my daily life when we were roommates. I must also thank my other collaborators, Dr. Irfan Alam, Dr. Sergii Myroshnychenko, Dr. Oscar Zatarain-Vera and Prof. Mark Meckes, for many useful mathematical discussions. I would also like to express my gratitude to Prof. Shiri Artstein-Avidan for many valuable comments on the initial draft of the second Chapter.

My PhD studies would have been incomplete without the wonderful memories in the classroom. I have benefited so much from the courses taught by the knowledgeable professors in the Mathematical department. In particular, I have learnt a lot in two classes: "Stochastic processes" and "Combinatorics", taught by Prof. Douglas Rizzolo and Prof. Felix Lazebnik, respectively. Especially, I would like to thank Prof. Felix Lazebnik for his encouragement. I would definitely also miss the good times outside the classroom that I spent in the company of my fellow graduate students: Melissa, Nick, Rayanne, Hansen, Yikun, Abhijit, Dheer & Taylor, Emma, Shashank, Hugo, Himanshu, Vishal, Avishek...Specifically, I would like to thank my current officemate Li-An Chen for her constant company since we met at the ELI, and I will always remember the afternoons when we hanged out together with Gautam on the veranda of the Graduate House. I hope that the friendship among us would last forever.

Lastly, this journey would not have been possible without the support of my family. Words cannot express my gratitude to my wife, Guoling Wei, who provides constant encouragement and has made countless sacrifices to help me get to this point. I would also like to express my gratitude to my parents for their continued support and unconditional love throughout my years of study.

To my wife, Wei Guoling

# TABLE OF CONTENTS

# ABSTRACT

The main theme of this dissertation is to explore four topics at the intersection of information theory, probability, discrete/convex geometry and geometric functional analysis.

In the first part of the dissertation, we introduce an information-theoretic approach to the Kneser-Poulsen conjecture in discrete geometry first formulated by Poulsen in 1954 and Kneser in 1955. Our approach revolves around a broad question regarding whether Rényi entropies of independent sums decrease when one of the summands is contracted by a 1-Lipschitz map. We answer this broad question affirmatively in various cases.

In the second part, we characterize the convex potentials, which is also known as the information content, of log-concave densities in $\mathbb{R}^n$ under various assumptions. Built upon this characterization, we show that, for a subclass of log-concave densities, the normalized information content satisfies both a central limit theorem and a large deviation principle, which, on one hand, generalizes the result for Gaussian random vectors, on the other hand, sheds some light on a more general conjecture.

In the third part, we investigate a metric generalization of Rényi entropies that are called diversities. We show that on the Euclidean metric space $\mathbb{R}^n$, one may recover the Rényi entropies from diversities. And via diversities, we define the diversity dimension of different orders of a Borel probability measure, and prove that they coincide with the information dimensions defined via Rényi entropies. Meanwhile, the fact that maximum diversity can also be used to recover some geometric invariants motivates us to generalize some classic sumset inequalities with maximum diversity now playing the role of "size". The relationship with the other notion of diversity is also explored in this section.

# Chapter 1

# INTRODUCTION

Information theory, founded by Claude Shannon in the 1940s, is a mathematical way of studying the communication of information. Despite its tight connections with communication theory, computer science and biology, the web of its tangled relationships with other abstract mathematical fields are elucidated in various previous works (see, e.g. [41], [75], [99], [61]). This dissertation explores the intersection of information theory with other mathematical topics such as probability, discrete/convex geometry and geometric functional analysis. In this chapter, we will briefly talk about three projects, from which the interplays between those topics will become apparent.

## 1.1 An information-theoretic approach to the Kneser-Poulsen conjecture in discrete geometry

If one starts with a finite number of open balls in a Euclidean space, then it appears plausible that the volume of their union should decrease if the centers are rearranged to be pairwise closer. This intuition was first formulated by Poulsen [89] and Kneser [60] independently.

**Conjecture 1** (Kneser–Poulsen)**.** Let $\{x_1, \ldots, x_k\}$ and $\{y_1, \cdots, y_k\}$ be two sets of points in $\mathbb{R}^d$ such that $\|y_i - y_j\|_2 \leq \|x_i - x_j\|_2$ for all $i, j \in \{1, \ldots, k\}$. Then for any $r > 0$, we have:

$$\mathrm{Vol}_d\left(\bigcup_{i=1}^{k} \mathcal{B}(y_i, r)\right) \leq \mathrm{Vol}_d\left(\bigcup_{i=1}^{k} \mathcal{B}(x_i, r)\right). \tag{1.1}$$

Bezdek and Connelly [11] proved this conjecture for union of open balls of arbitrary radii in the plane. So far, $d = 2$ remains the only dimension in which the conjecture has been proven completely. For an arbitrary Euclidean space $\mathbb{R}^d$, the conjecture has been partially proven under various additional assumptions. We refer the readers to recent surveys [33], [100] for a detailed account.

In our recent paper [1], we provided, to the best of our knowledge, a new approach to this conjecture using information theoretic ideas. Our approach is motivated by two observations.

First of all, the hypothesis in Conjecture 5 can be described in terms of a contraction defined on a subset of $\mathbb{R}^d$. Also, the union of open balls can be rephrased as the Minkowski sum (vector sum) of the set consisting of the centers and the open ball centered at the origin with radius $r$. Then Kirszbraun's extension theorem [55] enables us to extend this partially defined contraction to $\mathbb{R}^d$, which leads us to a reformulation of Conjecture 5:

**Conjecture 2.** For every contraction $T$ of $\mathbb{R}^d$ and every compact set $K \subseteq \mathbb{R}^d$, $r > 0$, we have

$$\mathrm{Vol}(T[K] + r\mathcal{B}) \leq \mathrm{Vol}(K + r\mathcal{B}),$$

where $\mathcal{B}$ denotes the open unit ball centered at the origin.

To prove a sumset volume inequality, one way is to prove a functional lifting of the volume inequality. For instance, the classical Brunn-Minkowski inequality (BMI) says that the volume of the convex combination of two compact subsets in $\mathbb{R}^d$ is no less than the geometric mean of the individual volumes:

$$\mathrm{Vol}_d((1-\lambda)A + \lambda B) \geq \mathrm{Vol}_d(A)^{1-\lambda}\mathrm{Vol}_d(B)^{\lambda},$$

for $\lambda \in [0, 1]$ and compact subsets $A$, $B$ in $\mathbb{R}^d$.

The functional lifting of the Brunn-Minkowski inequality is called the Prékopa-Leindler inequality: Let $\lambda \in (0, 1)$, and let $f$, $g$ and $h$ be nonnegative integrable functions on $\mathbb{R}^n$ satisfying $h((1-\lambda)x+\lambda y) \geq f^{1-\lambda}(x)g^{\lambda}(y)$ for all $x, y \in \mathbb{R}^d$. Then

$$\int_{\mathbb{R}^d} h(x)\ \mathrm{d}x \geq \left(\int_{\mathbb{R}^d} f(x)\ \mathrm{d}x\right)^{1-\lambda} \left(\int_{\mathbb{R}^d} g(x)\ \mathrm{d}x\right)^{\lambda}.$$

If one sets $f = \mathbf{1}_A$, $g = \mathbf{1}_B$, and $h = \mathbf{1}_{(1-\lambda)A+\lambda B}$, then the Brunn-Minkowski inequality follows from the Prékopa-Leindler inequality.

Actually, the Prékopa-Leindler inequality is a limit case of a more general inequality involving $L^p$ norms (when $p \in (0,1)$, the usual expression does not define a norm, but we will still call them norms), which is called the reverse Young's inequality. Young's inequality was first introduced by W. H. Young [105] in 1912 , then the reverse Young's inequality was found by Leindler [62] in 1972. Later, both Young's inequality and reverse Young's inequality with sharp constants were discovered by Beckner [9] in 1975. In the sequel, $f \star g$ represents the convolution between two functions $f$ and $g$.

**Theorem 1.1.1.** *Let $0 < p,\ q,\ r \leq 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, and let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$ be nonnegative. Then*

$$\|f \star g\|_r \geq C^d \|f\|_p \|g\|_q.$$

*Here $C = \frac{C_p C_q}{C_r}$, where*

$$C_s^2 = \frac{|s|^{1/s}}{|s'|^{1/s'}}$$

*for $1/s + 1/s' = 1$.*

Due to the implication from the reverse Young's inequality to Prékopa-Leindler inequality, one may expect a direct derivation of the Brunn-Minkowski inequality from the reverse Young's inequality. This was first observed by Dembo, Cover and Thomas [37] in 1991. We more or less follow this strategy by comparing the $L^p$ norms of convolutions of density function. To this end, we introduce the normalized $L^p$ norms, which are called Rényi entropies.

**Definition 1.** Let $X$ be an $\mathbb{R}^d$-valued random vector with density $f$ with respect to the Lebesgue measure. Then, the Rényi entropy of order $\alpha \in (0,1) \cup (1,\infty)$ of $X$ is given by,

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^\alpha \, \mathrm{d}x.$$

The Rényi entropy of orders $0, 1$, and $\infty$ are obtained via taking respective limits,

$$h_0(X) = \log \mathrm{Vol}(\mathrm{support}(f)),$$

$$h_1(X) = h(X) = -\int_{\mathbb{R}^d} f \log f \, \mathrm{d}x,$$

$$h_\infty(X) = -\log \|f\|_\infty.$$

The Rényi entropy of order 1 was called Shannon entropy, which was first introduced by Shannon [95] in 1948 , then extended by Rényi [91] in 1961. Loosely speaking, Rényi entropies measure the randomness of a random variable. Despite their appearance in data communication problems, Rényi entropies also show up in probability theory, functional analysis, additive combinatorics and convex geometry ( see, e.g. [104], [75], [99], [61]). For instance, apart from the functional lifting we mentioned earlier, there is also a probabilistic (entropic) lifting of the Brunn-Minkowski inequality which is called entropy power inequality. The EPI states that for any two independent continuous random vectors $X$ and $Y$ in $\mathbb{R}^d$ such that the Shannon entropies of $X$ and $Y$ and $X+Y$ exist, then

$$N(X+Y) \geq N(X) + N(Y),$$

where $N(X) = e^{\frac{2h_1(X)}{d}}$.

The entropy power inequality (EPI) was stated by Shannon [95] with an incomplete proof; the first complete proof was provided by Stam [98]. The EPI plays an extremely important role in the field of information theory, where it was used to prove statements about the fundamental limits of communication over various models of communication channels. Subsequently, it has also been recognized as an extremely useful inequality in probability theory, with close connections to the Log-Sobolev inequality [53] for the Gaussian distribution (see, e.g., [104]), as well as to the central limit theorem (see, e.g., [72]). The parallelism between the BMI and EPI was first observed by Costa and Cover [29]. Though they do not implies each other, they can be unified in at least two different ways: one way is via the reverse Young's inequality, which was explained by Dembo, Cover and Thomas [37]; the other way is by rearrangement technique, which goes back to Madiman and Wang [104].

Back to the conjecture, one observes that if $X$ is a random variable in $\mathbb{R}^d$ with density $f$ supported on a compact set $K$ of positive volume, then the Rényi entropy of order 0 recovers the volume.

These two observations hint that the Kneser–Poulsen conjecture can be further reinterpreted as a particular case of a broad information-theoretic question as follows:

**Question 1.** Let $X$ and $W$ be $\mathbb{R}^d$-valued random variables. Further assume that $W$ is log-concave and satisfies a symmetry property such as radial symmetry. For a contraction $T : \mathbb{R}^d \to \mathbb{R}^d$, and $\alpha \in [0, \infty]$, do we have

$$h_\alpha(T(X) + W) \leq h_\alpha(X + W)\,?$$

Indeed, if $K$ is a compact subset of $\mathbb{R}^d$, and $X$ is a random vector with support $K$, then for $W \sim \text{Uniform}(\mathcal{B})$, the support of $X + W$ is $K + \mathcal{B}$ and the support of $T(X) + W$ is $T[K] + \mathcal{B}$. Therefore, if the answer to Question 2 is true for all $\alpha$ sufficiently close to 0, then taking limits would yield the Conjecture 6 formulation of the Kneser–Poulsen conjecture.

Recall that a $\mathbb{R}^d$-valued random variable $X$ is said to be log-concave if its density function $f(x) = e^{-U(x)}$ for some convex function $U(x) : \mathbb{R}^d \to (-\infty, \infty]$. In our recent paper, we provided positive answers to Question 2 under various assumptions via different methods.

Firstly, note that when $W$ is radially symmetric log-concave, the density of $X+W$ always exists. We come up with a framework in the continuous setting which enables us to compare the "peakedness" of two different distributions given some extra geometric structure imposed on $X$. In particular, the comparison result implies the desired entropic inequality. One of our main results in this flavor is the following:

**Theorem 1.1.2.** *[1] If $X$ is a log-concave random vector and $W$ is a radially-symmetric log-concave random vector, then for any linear contraction $T$, we have*

$$h_\alpha(T(X)+W) \leq h_\alpha(X+W) \text{ for all } \alpha \in (0,\infty).$$

The key inequality that we prove in the theorem is the following:,

**Theorem 1.1.3.** *[1] For any convex body $K$ in $\mathbb{R}^d$, any diagonal matrix $T$ with diagonal entries in $[0,1]$, and $g(x)$ unconditional log-concave or radially symmetric unimodal, there exists a convex body $S(K)$ having same volume as $K$, such that*

$$\int_{S(K)} g(x+T(y))\, \mathrm{d}x \geq \int_K g(x+y)\, \mathrm{d}x$$

*for all $y \in \mathbb{R}^d$.*

On one hand, Theorem 1.1.3 is reminiscent of an inequality proved by Anderson [4] in 1955, which requires $K$ to be origin symmetric and $T$ be a multiple of the identity matrix, but with weaker condition on $g(x)$. On the other hand, the technique used in proving this theorem provides a different proof of the fact that intrinsic volumes decrease under linear contractions, which is a result credited to Paouris and Pivovarov [86].

Secondly, we explored the special case when $W$ is isotropic Gaussian and $\alpha = 1$. With an aid of the vector generalization of Costa's EPI due to Liu, Liu, Poor and Shamai [71] and Courtade, Han and Wu [30], we show that in this special case, the entropic inequality holds for linear contraction $T$. Moreover, using the classic moment-entropy inequality combined with variance estimation, we show positive results when

$X$ is Gaussian whose covariance matrix is diagonal and $T$ is a strong contraction (coordinatewise, $T$ is also a contraction). The results in this section leads us to a natural conjecture which we state below

**Conjecture 3.** Let $Z \sim \mathcal{N}(0, I_d)$ be the standard Gaussian random vector in $\mathbb{R}^d$, and $X$ be an independent random vector in $\mathbb{R}^d$ with density, and $T$ be any contraction. Then we have

$$N(X+Z) \geq N(T(X)+Z) + (1 - \mathrm{Lip}^2(T))N(X).$$

*Remark.* We show that the conjecture is true when $T$ is a linear contraction. In particular, when $T(x) = \lambda x,\ \lambda \in [0,1]$, we recover Costa's EPI [28].

At the end of this chapter, using a metric generalization of Rényi entropies, which are called diversities, we give an affirmative answer to Question 2 when $\alpha = 2$. Specifically, we show that

**Theorem 1.1.4.** *[1] Let $X$ be any $\mathbb{R}^d$-valued random variable and $W$ be any $\mathbb{R}^d$-valued radially-symmetric log-concave random variable, then for any contraction $T$, we have*

$$h_2(T(X)+W) \leq h_2(X+W).$$

Though the Conjecture 5 has not been resolved in our work, we consider this novel approach to be the first steps toward an information-theoretic interpretation of Kneser–Poulsen-type questions in metric geometry. We hope that this connection leads to developments that enrich both fields. For example, there is a straightforward connection between channel capacity of a certain additive noise channel and the Kneser-Poulsen conjecture, which will be discussed in the last section.

## 1.2 A central limit theorem of information content for log-concave densities

Consider the standard Gaussian random vector $Z = (Z_1, \ldots, Z_n)$ in $\mathbb{R}^n$, the density of which is given by

$$\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{2}}. \tag{1.2}$$

It follows from the weak law of large numbers that $\frac{\|Z\|^2}{n}$ converges to 1 in probability as the dimension $n$ goes to infinity. In other words, for any fixed $t > 0$ and $\epsilon > 0$, when $n$ is sufficiently large, one has

$$\mathbb{P}\left\{ \frac{\|Z\|^2}{n} - 1 > t \right\} < \epsilon. \tag{1.3}$$

There are some well known ways of quantifying the inequality (1.3). One of them is to apply the Markov inequality to the function $e^{\lambda x}$ for some suitably chosen $\lambda$, which yields the so-called Chernoff's bound:

$$\mathbb{P}\left\{ \frac{\|Z\|^2}{n} - 1 > t \right\} \leq \exp\left\{ -\frac{n}{2}[t - \log(1+t)] \right\}. \tag{1.4}$$

The bound is asymptotically sharp in the sense that one has

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{ \frac{\sum_{i=1}^n Z_i^2}{n} - 1 > t \right\} = \frac{\log(1+t) - t}{2}, \tag{1.5}$$

which can be viewed as a large deviation principle (see, e.g., [36]) for the laws of $\frac{\|Z\|^2 - n}{n}$.

Another way to quantify the inequality (1.3) is by noticing that $\|Z\|^2$ is a sum of i.i.d. random variables, so that one may use the Berry-Esseen theorem to obtain

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\|Z\|^2 - n}{\sqrt{2n}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \, \mathrm{d}r \right| \leq \frac{c}{\sqrt{n}}, \tag{1.6}$$

where $c$ is an absolute constant.

Recall that log-concave probability densities are densities that can be written as $f(x) = e^{-U(x)}$, where $U(x)$ is a convex function defined on $\mathbb{R}^n$. They can be thought of as a natural generalization of Gaussian densities. Therefore, it is natural to wonder if the above properties of Gaussian densities can be inherited by log-concave densities. Toward that end, notice that $\|Z\|^2 - n = 2[-\log \phi(Z) + \mathbb{E} \log \phi(Z)]$. Thus, for a random vector $X$ in $\mathbb{R}^n$ with log-concave density $f$, it is natural to consider the quantity $-\log f(X) + \mathbb{E} \log f(X)$ . However, we encounter an issue: due to the lack of independence, we are unable to get an inequality like (1.3) for the random variable

$$\frac{-\log f(X) + \mathbb{E} \log f(X)}{n} \tag{1.7}$$

via the weak law of large numbers. However, the following sharp bound (1.8) for the variance of the random variable $-\log f(X)$ turns everything around:

$$V(X) := \mathrm{Var}(-\log f(X)) \leq n. \tag{1.8}$$

This inequality was proved independently by Nguyen [85] and Wang [103], then sharpened by Bolley, Gentil and Guillin [19]. Using it, a direct application of Chebyshev's inequality to the random variable in (1.7) enables us to derive a result like inequality (1.3).

Later, Fradelizi, Madiman, and Wang [43] came up with an easier proof of the variance bound (1.8), and attained the following optimal concentration inequality for the random variable $-\log f(X)$, which generalizes the inequality (1.4):

$$\mathbb{P}\left\{-\log f(X) + \mathbb{E} \log f(X) > nt\right\} \leq \exp\left\{-nr(t)\right\}, \tag{1.9}$$

$$\mathbb{P}\left\{-\log f(X) + \mathbb{E} \log f(X) < -nt\right\} \leq \exp\left\{-nr(-t)\right\}, \tag{1.10}$$

where $r(t) = t - \log(1+t)$ for $t > -1$, and $r(t) = +\infty$, otherwise.

As a consequence of inequality 1.10, one deduces the following entropy comparison inequality, which was first explicitly proved in [15],

$$\mathbb{E}[-\log f(X)] \leq -\log \|f\|_\infty + n. \tag{1.11}$$

In this dissertation, we characterize the equality case in (1.11), which in turn implies that inequality (1.9) is asymptotically sharp when the entropy of $X$ is saturated, in the same way that Equation (1.4) is asymptotically sharp as indicated by Equation (1.5). Moreover, we observe that all these results are actually special cases of a more general result. For more details, an interested reader is referred to the Chapter 3 of this dissertation.

Another consequence of the concentration inequalities is the following result [43]: Let $X$ be a log-concave random variable in $\mathbb{R}^n$ with a log-concave density $f(x)$. Then

$$\mathbb{P}\{f(X) \geq c^n \|f\|_\infty\} \geq 1 - \left(e \cdot c \cdot \log\left(\frac{1}{c}\right)\right)^n,$$

where $0 < c < \frac{1}{e}$. When $f$ is a log-concave density, the convex body defined by

$$K_f = \{x \in \mathbb{R}^n : f(x) \geq c^n \|f\|_\infty\}$$

may be viewed as the "effective support" of the law of $X$. The name "effective support" can be justified by choosing a suitable value of $c$ such that the right hand side is close to 1. Such "effective support" results are useful in convex geometry as they can be used to reduce certain statements about log-concave functions or measures to statement about convex set; they thus provide an efficient route to proving functional or probabilistic analogues of known results in convex geometry. For instance, the reverse Brunn-Minkowski inequality is a deep result in convex geometry discovered by V.D. Milman [83]. It states that, given two convex bodies $A$ and $B$ in $\mathbb{R}^n$, one can find a linear volume preserving map $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that with some absolute constant

$C$,

$$\mathrm{Vol}_n(T(A)+B)^{1/n} \leq C\left(\mathrm{Vol}_n(A)^{1/n}+\mathrm{Vol}_n(B)^{1/n}\right)$$

This remarkable result in convex geometry also enjoys a functional lifting to log-concave functions [58] and a probabilistic (entropic) lifting to log-concave random variables [16] (actually, the result in [16] holds in more general setting). Interestingly, the "effective support" was employed in both of the proofs of these lifting results. For other applications of the concentration of information phenomenon, interested readers are referred to the paper [41].

The ubiquitous random variable $-\log f(X)$ is called the *information content* of the random vector $X$. In one of our forthcoming papers, we show that for a subclass of log-concave densities, the properly normalized information content satisfies both a large deviation principle and a central limit theorem just like (1.5) and (1.6). Specifically, we prove the following theorem.

**Theorem 1.2.1.** *[68] Let $\beta > 0$ be a constant and $m$ be a positive integer, and suppose that $\mathbb{R}^n = E_1 \oplus E_2 \oplus \ldots \oplus E_m$ with dim$(E_i) = k_i$. Let $\gamma = (\gamma_1, \ldots, \gamma_m) \in [1, \infty)^m$ be such that*

$$\sum_{i=1}^{m} \frac{k_i}{\gamma_i} = \beta n. \tag{1.12}$$

*Also, let $\|\cdot\|_{E_i}$ be any (asymmetric) seminorm in each $E_i$ and choose a positive constant $C$ such that*

$$f(x) = Ce^{-\sum_{i=1}^{m} \|P_{E_i}x\|_{E_i}^{\gamma_i}} \tag{1.13}$$

*is a probability density, where $P_{E_i}$ denotes the orthogonal projection onto subspace $E_i$. Then for any log-concave random vector $X$ in $\mathbb{R}^n$ with density $f(x)$,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{-\log(f(X)) + \mathbb{E}\log(f(X))}{\sqrt{V(X)}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}}\, \mathrm{d}r \right| \leq \frac{c}{\sqrt{n}},$$

*where $c$ is an absolute constant that only depends on the positive number $\beta$.*

In the special case when $m = n, \gamma_i = 2$ for all $i$, and $\beta = \frac{1}{2}$, we immediately recover the inequality (1.6) if we apply Theorem 1.2.1 to the standard Gaussian density $f$ defined in (1.2). However, since arbitrary norms are allowed in (1.13), the scope of the theorem is much wider than just the Gaussian densities. Moreover, since the constant $c$ in the theorem's conclusion does not depend on the choice of the positive integer $m$ and vector $\gamma$, this potentially opens up further research directions. In particular, we hope that our work will bring the readers' attention to the following more general conjecture formulated in our paper [68].

**Conjecture 4.** Given a random vector $X$ in $\mathbb{R}^n$ with log-concave density $f(x) = e^{-U(x)}$, suppose that

$$\mathrm{Var}(-\log f(X)) = \beta n, \ \beta > 0,$$

then,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{U(X) - \mathbb{E}[U(X)]}{\sqrt{V(X)}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}}\, \mathrm{d}r \right| \leq \frac{c}{\sqrt{n}},$$

where $c$ is an absolute constant that only depends on $\beta$.

It is worth mentioning that, in contrast to the central limit theorem established by B. Klartag [56], [57], which asserts that given any isotropic log-concave random vectors $X$, for most of the directions $\theta \in S^{n-1}$, the total variation distance between $\langle X, \theta \rangle$ and the standard normal random variable $\mathcal{N}(0,1)$ is bounded by $1/n^k$ for some universal constant $k > 0$, our attention is focused on the normalized random variable

$$\frac{-\log f(X) + \mathbb{E}(\log f(X))}{\sqrt{V(X)}},$$

which is affine-invariant (see details in Section 3.2). Since any log-concave random vector can be placed in an isotopic position via an affine transformation (see, e.g., [22, Chapter 2]), we may also assume that $X$ is isotropic. The proof of Klartag's central limit theorem boils down to the proof of the so called thin shell condition (see, e.g., Bobkov [17]); while our qualitative central limit theorem Conjecture 9 can be fully settled if one is able to prove a uniform bound for the third moments of a family of random variables (see Theorem 3.3.3 for more details).

### 1.3 Diversity and dimension of probability measures in metric spaces

Let $X$ be a discrete random variable with possible outcomes in a finite subset of $\mathbb{R}^d$, say, $\mathcal{S} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ and corresponding probabilities $p_i = \mathbb{P}\{X = x_1\}$ for $i = 1, \ldots, n$. The Rényi entropies of order $\alpha \in [0,1) \cup (1,\infty)$ is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left( \sum_{i=1}^{n} p_i^\alpha \right). \tag{1.14}$$

The Rényi entropy of orders 1 and $\infty$ are obtained via taking respective limits,

$$H_1(X) = -\sum_{i=1}^{n} p_i \log p_i, \quad H_\infty(X) = -\log \max_i p_i.$$

From the definition, one realizes that Rényi entropies of different orders tell us information about the distribution of the random variable $X$. For instance, Rényi entropy of order 0 tells us the cardinality of all possible outcomes, while Rényi entropy of order $\infty$ tells us the information of the most likely outcome(s). Moreover, one has the following inequality for Rényi entropies of all orders,

$$0 \leq H_\alpha(X) \leq \log n \;\text{ for all } \alpha \in [0,\infty].$$

It can be shown that for all $\alpha$, the lower bound is attained if and only if $X$ is supported on a singleton, while the upper bound is achieved if and only if $X$ has the uniform distribution on $\mathcal{S}$. Therefore, in some sense, Rényi entropies measure the expected uncertainty of a random variable $X$.

However, one may immediately realize a shortcoming of Rényi entropies. Consider another discrete random variable $Y$ with possible outcomes in the set $\mathcal{B} = \{y_1, y_2, \ldots, y_n\} \subset \mathbb{R}^n$, and assume that $q_i = \mathbb{P}\{Y = y_i\} = \mathbb{P}\{X = x_i\} = p_i$. Then one always has $H_\alpha(X) = H_\alpha(Y)$. In other words, Rényi entropies only tell us the information about the distributions over sets, but fail to detect any information about the geometric structures of those sets, which appears crucial in both abstract theories and concrete applications. Therefore, it is natural to come up with generalizations that take both the distribution of $X$ and the metric structure of $\mathcal{S}$ into consideration. *Diversity* is one such generalization [67].

**Definition 2.** Consider the discrete random variable $X$ taking values in a finite subset $\mathcal{S}$ in $\mathbb{R}^d$, say, $\mathcal{S} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$, and equip $\mathbb{R}^n$ with an arbitrary metric $d$. Then the diversity of order $\alpha \in (-\infty, 1) \cup (1, \infty)$ is defined as:

$$D_\alpha^{\mathcal{S}}(p) = \left( \sum_i \left( \sum_j e^{-d(x_i, x_j)} p_j \right)^{\alpha - 1} p_i \right)^{\frac{1}{1-\alpha}}. \tag{1.15}$$

For $\alpha = \pm\infty, 1$, we take the limiting values of the above expression.

In a straightforward manner, one can generalize the definition to an arbitrary metric space $(\mathcal{S}, d)$ and any Borel probability measure $\mu$ on $\mathcal{S}$:

$$D_\alpha^{\mathcal{S}}(\mu) = \left( \int \left( \int e^{-d(x,y)} d\mu(y) \right)^{\alpha - 1} d\mu(x) \right)^{\frac{1}{1-\alpha}}. \tag{1.16}$$

It turns out that this metric generalization of Rényi entropy is quite powerful. Firstly, from a geometric point of view, the maximization theorem in [66] allows one to define the *maximum diversity* $|\mathcal{S}|_+$ of a compact metric space $\mathcal{S}$, which, in turn, encodes several interesting invariants in integral geometry and geometric measure theory (see also [65]). Such invariants include the Minkowski dimensions of compact metric spaces. Meanwhile, the variational approach of maximum diversity via reproducing kernel Hilbert spaces also unveils its tight connection with capacity theory. The interested readers are referred to [80], [65] for more information. In this thesis, new connections with other areas are explored. In particular, we reveal and exploit the relationships between diversity, information theory, convex geometry, and optimal transport. We briefly sketch some of the themes of this section next.

Before moving forward, we would like to point out that, in fact, diversities can be defined in any Hausdorff space $X$. Let $X$ be a Hausdorff space, and let $K(x, y) : X \times X \mapsto [0, \infty)$ be a symmetric continuous function satisfying $K(x, x) > 0$ for all $x \in X$. The function $K$ is called a similarity kernel. For points $x$ and $y$ in $X$, one may think of them as species in an ecosystem, and $K(x, y)$ as similarity between these two species. For a finite space $X$, Cobbold and Leinster believe that realistic measures of biodiversity should reflect not only the relative abundances of species (the distribution), but also the differences between them (the similarities). Hence, in [67], they introduced an one parameter family of entropy measures as follows:

$$D_\alpha^K(p) = \left( \sum_i \left( \sum_j K(i, j) p_j \right)^{\alpha - 1} p_i \right)^{\frac{1}{1-\alpha}}$$

Then for an ecosystem with $n$ species (the similarity between each species are therefore aware), one would like to know if, by suitably choosing the relative abundances of species, we can maximize the biodiversity of the ecosystem. In other words, one would like to compute

$$\max_p D_\alpha^K(p)$$

A compactness argument implies that the maximizer always exists for a compact space $X$ and fixed $\alpha \geq 0$. However, a maximization theorem due to Leinster and Meckes [64] for any finite space, then generalized to any compact Hausdorff space by Leinster and Roff [66], says that there always exists a probability measure $\mu$ on $X$ that maximizes $D_\alpha^K(\mu)$ for all $\alpha \geq 0$ simultaneously. The maximization theorem has nice applications in both graph theory and metric geometry (see, e.g., [64]).

For a finite metric space $X$, the function $e^{-d(x,y)}$ naturally serves as a similarity kernel. Replacing the metric $d$ with the metric $td$, then sending $t$ to $\infty$, one easily recovers the Rényi entropies in a finite metric space. It might not be too surprising that same technique can be used to recover Rényi entropies in an Euclidean space $\mathbb{R}^d$, as well. Moreove, just like Minkowski dimension serves as a measure of the fractal dimension of a set, *information dimension* can be thought of as a measure of the fractal dimension of a probability distribution. Information dimension (of order 1) was first introduced by Alfréd Rényi [90] in 1959 via Rényi entropy of order 1. It was then generalized to different orders by Csiszár [34]. What Rényi and Csiszár did consists of two steps. First, using discretizations, approximate the given probability distribution $\mu$ by a sequence of discrete distributions. Then the Rényi entropies (of order $\alpha$) of these discrete distributions are scaled by suitable factors. If the limit of these scaled Rényi entropies exists, then it is called the information dimension of $\mu$ (of order $\alpha$). Note that the discretization process itself is implicitly connected with the metric structure of the space. Therefore, one may wonder if the information dimension can be recovered directly from diversity via scaling the metric of the space. This curiosity drove us to analogously define a notion of *diversity dimension* in our forthcoming paper [2], using which we were indeed able to recover the information dimension.

**Theorem 1.3.1.** *[2] Let $\mu$ be a Borel probability distribution on $\mathbb{R}^n$. For any $\alpha > 1$, the information dimension of $\mu$ of order $\alpha$ exists if and only if the diversity dimension of $\mu$ of order $\alpha$ exists. In that case, they are equal. If $\mu$ is compactly supported, then the same conclusion holds for $\alpha \in (0, 1)$.*

A metric space is called positive definite, if the kernel function $e^{-d(x,y)}$ is positive definite. In view of the intimate relationship between maximum diversity and metric/integral geometry indicated in [80], [65] for positive definite metric spaces, we show the fractional subadditivity property of maximum diversity when the underlying metric space is positive definite.

**Theorem 1.3.2.** *[2] Let $(X, d)$ be a positive definite metric space, $A_1, A_2, \cdots, A_n \subseteq X$ be compact subsets, and $\beta : 2^{[n]} \to [0, \infty)$ be a fractional partition function. Then*

$$\left|\bigcup_{i=1}^{n} A_i\right|_+ \leq \sum_{s \in 2^{[n]}} \beta(s) \left|\bigcup_{i \in s} A_i\right|_+ .$$

Meanwhile, we are also interested in understanding when classic sumset inequalities (both for cardinalities and volumes) can be generalized with maximum diversity now playing the role of "size". In [2], we also obtain some results of this nature. One such result which resembles the classic Brunn–Minkowski inequality is listed below.

**Theorem 1.3.3.** *[2] Let $A$ and $B$ be two compact sets in $\mathbb{R}^n$ (equipped with the usual Euclidean metric) and $0 < \lambda < 1$. Then there exists an absolute constant $c \in (0,1)$ such that*

$$|(1-\lambda)A + \lambda B|_+ \geq \begin{cases} |A|_+^{1-\lambda}|B|_+^{\lambda}, & \text{if } n = 1, \\ |cA|_+^{1-\lambda}|cB|_+^{\lambda}, & \text{if } n \geq 2. \end{cases}$$

In the end of this section, we explore the relationship with the notion of diversity introduced by Bryant and Tupper [25] (BT diversity):

**Definition 3.** Let $X$ be a a set, $\delta : \{F \in 2^X : \#F < \infty\} \to [0, \infty)$. The pair $(X, \delta)$ is called a diversity if $\delta$ satisfies the conditions:

1. $\delta(A) = 0$ if and only if $\#A \leq 1$;
2. $\delta(A \cup B) \leq \delta(A \cup C) + \delta(C \cup B)$, whenever $B \neq \emptyset$.

If we consider the function $\delta(A) = \max\{|A|_+ - 1, 0\}$, then we have the following result:

**Theorem 1.3.4.** *[2] Suppose $(X, d)$ is a metric space such that all non-empty finite subsets are viewed as pointed metric spaces. Then, the corresponding $(X, \delta)$ is a BT diversity.*

In the first part of the introduction, we noticed that this metric generalization of Rényi entropies might be useful in proving some probabilistic inequality where metric is involved. In the last section of the dissertation, we will discuss a possible connection between diversity and the Kneser-Poulsen conjecture. Also, a concavity conjecture of diversity of order 1 in a finite space where the similarity kernel function is positive definite will be discussed. The conjecture arises from the fact that it is known to be true for Rényi entropy of order 1.

# Chapter 2

# AN INFORMATION-THEORETIC APPROACH TO KNESER-POULSEN CONJECTURE

## 2.1 Background

As discussed in the introduction, the Kneser–Poulsen conjecture in discrete geometry asserts that when the centers of a finite number of open balls in a Euclidean space are brought closer, one would expect that the volume of the union of the balls decrease. Before stating this conjecture formally, let us fix some notation.

*Notation.* Throughout this section, we will work in the Euclidean space $\mathbb{R}^d$, where $d \in \mathbb{Z}_{>0}$. The $L^2$-norm on $\mathbb{R}^d$ is denoted by $\|\cdot\|_2$, where we will often suppress the subscript and write $\|\cdot\|$ when there is no risk of confusion. Throughout the paper, the metric on $\mathbb{R}^d$ that is used (for instance, in describing Lipschitz functions) is the one induced by $\|\cdot\|$. We use $\mathrm{Vol}_d$ (or just Vol when there is no risk of confusion) to denote the Lebesgue measure on $\mathbb{R}^d$. The open ball of radius $r$ centred at the point $x \in \mathbb{R}^d$ will be denoted by $\mathcal{B}(x, r)$, while the ball $\mathcal{B}(0, 1)$ will be denoted simply by $\mathcal{B}$.

**Conjecture 5** (Kneser–Poulsen)**.** Let $\{x_1, \ldots, x_k\}$ and $\{y_1, \cdots, y_k\}$ be two sets of points in $\mathbb{R}^d$ such that $\|y_i - y_j\|_2 \le \|x_i - x_j\|_2$ for all $i, j \in \{1, \ldots, k\}$. If $r > 0$, then we have:

$$\mathrm{Vol}_d\left(\bigcup_{i=1}^{k} \mathcal{B}(y_i, r)\right) \le \mathrm{Vol}_d\left(\bigcup_{i=1}^{k} \mathcal{B}(x_i, r)\right). \tag{2.1}$$

Bezdek and Connelly [11] proved this conjecture in the plane. So far, $d = 2$ remains the only dimension in which the conjecture has been proven completely. For an arbitrary Euclidean space $\mathbb{R}^d$, the conjecture has been proven under additional assumptions on the number of points $k$ and the map $x_i \mapsto y_i$. For example, Csikós [32] proved the Kneser–Poulsen conjecture under the assumption that each initial point $x_i$ can be joined with the corresponding final point $y_i$ by a path such that all pairwise distances decrease along the path. Such a requirement is not always satisfied if the number of points involved exceeds the dimension. More recently, Bezdek and Naszódi [12] demonstrated the conjecture for uniform contractions, that is, when there exists $\lambda > 0$ such that $\|y_i - y_j\|_2 < \lambda < \|x_i - x_j\|_2$ for all $i \neq j$. In the same work, the authors also settle the case when the pairwise distances are reduced in every coordinate. In this quick literature review, we have skipped many rich developments around the conjecture including those pertaining to its formulation in other spaces. We refer the reader to the book [10], or recent surveys [33, 100] for a more detailed account.

The results in our paper are motivated by our attempt to understand the Kneser–Poulsen conjecture from an information-theoretic viewpoint. We begin by gathering the necessary background to formulate our results.

Let $S$ denote the set $\{x_1, \cdots, x_k\}$ and let $T\colon \mathbb{R}^d \to \mathbb{R}^d$ be a 1-Lipschitz map (which we will call a *contraction*) such that

$$y_i = T(x_i) \text{ for all } i \in \{1, \dots, d\}. \tag{2.2}$$

Given the points $\{y_1, \dots, y_d\}$ as in the statement of Conjecture 5, a contraction $T$ satisfying (2.2) indeed exists by Kirzbraun's theorem [55]. For a fixed $r > 0$, observe that the set $\bigcup_{i=1}^k \mathcal{B}(x_i, r)$ can be rewritten as $S + r\mathcal{B}$. Using an elementary approximation-from-within argument to go from finite sets to arbitrary compact sets, we can thus rephrase the Kneser-Poulsen conjecture in the following form.

**Conjecture 6.** For every contraction $T$ of $\mathbb{R}^d$ and every compact set $K \subseteq \mathbb{R}^d$, $r > 0$, we have

$$\mathrm{Vol}(T[K] + r\mathcal{B}) \leq \mathrm{Vol}(K + r\mathcal{B}).$$

This reformulation of the Kneser–Poulsen conjecture can be further reinterpreted as a particular case of a broad information-theoretic question. The dissertation studies that information-theoretic question, for which we manage to give positive answers to several cases.

To the best of our knowledge, this method is new in the literature of the Knesen–Poulsen conjecture. We consider this novel approach to be the first steps towards an information-theoretic interpretation of Kneser–Poulsen-type questions in metric geometry. We hope that this connection leads to developments that enrich both fields. For example, our results have straightforward implications for channel capacities of certain additive noise channels. These implications and a channel capacity based approach to Conjecture 6 will be presented in a follow up note.

Let us begin by establishing some notation needed to express the questions that we aim to study.

### 2.1.1 Some preliminary notation and definitions

Throughout, unless stated otherwise, all sums $X + Y$ that we study will be sums of independent random vectors. The distribution of a random vector $X$ will sometimes be denoted by $\mathbb{P}_X$, and if it happens to have a density with respect to the Lebesgue measure on the ambient space, its density will sometimes be denoted by $f_X$. Recall the definition of Rényi entropies: Let $X$ be an $\mathbb{R}^d$-valued random vector with density $f$ with respect to the Lebesgue measure. Then, the Rényi entropy of order $\alpha \in (0,1) \cup (1,\infty)$ of $X$ is given by,

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^\alpha(x) \, \mathrm{d}x.$$

The Rényi entropy of orders $0, 1$, and $\infty$ are obtained via taking limits. For an $\mathbb{R}^d$-valued random vector $X$ with distribution $\mu \in \mathcal{P}(\mathbb{R}^d)$ ($\mathcal{P}(\mathbb{R}^d)$ denotes the space of Borel probability measures) having density $f$ with respect to the Lebesgue measure, we will abuse notation and use $h_\alpha(X), h_\alpha(\mu)$, and $h(f)$, interchangeably.

It has been realized through a continuous stream of works in the past few decades that entropy can be powerfully used as a proxy for volume (and other geometric invariants) to bring concrete geometric problems to the setting of information theory and probability where it becomes susceptible to more analytic tools. This technique has been especially fruitful in the subject of convex geometry (see [75], and the references therein). Information-theoretic analysis is often more effective when the underlying random variables have nice geometric structure. Unsurprisingly, such structures have a convexity flavor. Some, which make an appearance in our results, are defined below.

**Definition 4.** Let $X$ be an $\mathbb{R}^d$-valued random vector with density $f$ with respect to the Lebesgue measure.

- $X$ is said to be log-concave if $f = e^{-\phi}$, for some convex function $\phi : \mathbb{R}^d \to (-\infty, \infty]$.
- $X$ is said to be unconditional if its distribution is invariant under reflections about the coordinate axes.
- $X$ is unimodal if $\{f > t\}$ is convex for every $t \in \mathbb{R}$. The terminology "unimodal" stems from the fact that the density of a unimodal random vector has a single "peak".
- $X$ is said to be isotropic if its covariance matrix is a scalar multiple of the identity matrix.

We use this opportunity to also define the special classes of contractions that we use in this paper.

**Definition 5.** Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a contraction.

- $T$ is an affine contraction if $T(x) = A(x) + b$, where $A$ is a linear map and $b \in \mathbb{R}^d$ a fixed vector. Such a $T$ is a linear contraction if $b = 0$.
- $T$ is a diagonally linear contraction if $T$ is given by $T(x_1, \dots, x_d) = (\lambda_1 x_1, \dots, \lambda_d x_d)$ for some fixed $\lambda_1, \dots, \lambda_d \in \mathbb{R}$. Since $T$ is already assumed to be a contraction, it follows that each $|\lambda_i| \leq 1$.
- $T$ is said to be a strong contraction if $T = (T_1, \dots, T_d)$ and $|T_i(x) - T_i(y)| \leq |x_i - y_i|$ for all $i = 1, \dots, d$ and all $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathbb{R}^d$.

### 2.1.2 Our information-theoretic question and Kneser–Poulsen-type problems

Consider the following slightly open-ended question.

**Question 2.** Let $X$ and $W$ be $\mathbb{R}^d$-valued random vectors. Further assume that $W$ is log-concave and satisfies a symmetry property such as radial symmetry, or unconditionality, etc. For a contraction $T : \mathbb{R}^d \to \mathbb{R}^d$, and $\alpha \in [0, \infty]$, under what additional assumptions do we have

$$h_\alpha(T(X) + W) \leq h_\alpha(X + W)\,?$$

Let $K$ be a compact subset of $\mathbb{R}^d$. Let $X$ be a random vector with support $K$ and let $W \sim \text{Uniform}(\mathcal{B})$. Then the support of $X + W$ is $K + \mathcal{B}$ and the support of $T(X) + W$ is $T[K] + \mathcal{B}$. If the answer to Question 2 is true for all $\alpha$ sufficiently close to 0, then taking limits yields the Conjecture 6 formulation of the Kneser–Poulsen conjecture.

Let $\alpha = n \geq 2$ be an integer and $W \sim \text{Uniform}(\mathcal{B})$. Suppose $X$ is a discrete random vector taking the values $x_i$ with probability $p_i$, $i = 1, \ldots, k$, respectively. Then $X+W$ has density $\frac{1}{\text{Vol}(\mathcal{B})} \sum p_i \mathbf{1}_{\mathcal{B}}(x-x_i)$. Similarly, if $y_i = T(x_i)$ (where $T\colon \mathbb{R}^d \to \mathbb{R}^d$ is a contraction, as before), then the random vector $T(X)+W$ has density $\frac{1}{\text{Vol}(\mathcal{B})} \sum p_i \mathbf{1}_{\mathcal{B}}(x-y_i)$. In this setup, Question 2 takes the following form

$$\int_{\mathbb{R}^d} \left( \frac{1}{\text{Vol}(\mathcal{B})} \sum_{i=1}^{k} p_i \mathbf{1}_{\mathcal{B}}(x - y_i) \right)^n \, \mathrm{d}x \geq \int_{\mathbb{R}^d} \left( \frac{1}{\text{Vol}(\mathcal{B})} \sum_{i=1}^{k} p_i \mathbf{1}_{\mathcal{B}}(x - x_i) \right)^n \, \mathrm{d}x\,?$$

If one first expands the integrands using the multinomial theorem and then proceeds to a term-by-term comparison, one realizes that an affirmative answer can be obtained if

$$\text{Vol}\left( \bigcap_{i \in S} \mathcal{B}(y_i, r) \right) \geq \text{Vol}\left( \bigcap_{i \in S} \mathcal{B}(x_i, r) \right),$$

for every $S \subseteq \{1, \ldots, k\}$ of cardinality at most $n$. This brings us to a dual formulation of Conjecture 5, first investigated by Gromov [52] and by Klee and Wagon [59].

**Conjecture 7.** Let $\{x_1, \ldots, x_k\}$ and $\{y_1, \cdots, y_k\}$ be two sets of points in $\mathbb{R}^d$ such that $\|y_i - y_j\|_2 \leq \|x_i - x_j\|_2$ for all $i, j \in \{1, \ldots, k\}$. For $r > 0$, we have:

$$\mathrm{Vol}\left(\bigcap_{i=1}^{k} \mathcal{B}(y_i, r)\right) \geq \mathrm{Vol}\left(\bigcap_{i=1}^{k} \mathcal{B}(x_i, r)\right). \tag{2.3}$$

The previous considerations with the multinomial theorem show that the affirmative answer to Conjecture 7 when the number of points involved is at most $k$, implies the desired Rényi entropic comparisons for integer orders $\alpha = 2, \cdots, k$.

**Proposition 2.1.1.** *The intersection version of the Kneser-Poulsen conjecture (i.e. Conjecture 7) implies,*

$$h_k(T(X) + W) \leq h_k(X + W),$$

*for any $\mathbb{R}^d$-valued random vector $X$, where $W \sim \mathrm{Uniform}(\mathcal{B})$ and $k \geq 2$ is an integer.*

Starting from Gromov's work [52] which established it for at most $d+1$ balls, Conjecture 7 is now known for at most $d+3$ balls in $\mathbb{R}^d$ by the work of Bezdek and Connelly [11, Corollary 4]. Thus, we have the following corollary.

**Corollary 2.1.2.** *Let $X$ be an $\mathbb{R}^d$-valued random vector, $T : \mathbb{R}^d \to \mathbb{R}^d$ any contraction and $W \sim \mathrm{Uniform}(\mathcal{B})$. Then for $\alpha = 2, 3, \ldots, d+3$, we have*

$$h_\alpha(T(X) + W) \leq h_\alpha(X + W).$$

Returning to the entropic analogues of the (union) Kneser-Poulsen conjecture, we will first try to answer this question using a systematic method.

## 2.2 Rearrangement methods

We will now rearrange the density of random variables $X + W$ and $T(X) + W$ into radially-symmetric unimodal densities while keeping their Rényi entropies fixed.

As it will be evident, comparing Rényi entropies of two radially-symmetric unimodal densities is easier than the general case.

**Definition 6.** For every Borel set $A \subseteq \mathbb{R}^d$ of positive volume, let $A^*$ denote the centered Euclidean ball in $\mathbb{R}^d$ having the same volume as $A$. Then, for a non-negative measurable $f : \mathbb{R}^d \to [0, \infty)$ which vanishes at infinity, we define its symmetrically-decreasing rearrangement as an almost-everywhere uniquely defined function $f^* : \mathbb{R}^d \to [0, \infty)$ characterized by the property

$$\{x : f^*(x) > t\} = \{x : f(x) > t\}^*,$$

for every $t > 0$.

*Remarks.*

1. One can describe $f^*$ explicitly by formula
$$f^*(x) = \int_0^\infty \mathbf{1}_{\{y: f(y) > t\}^*}(x) \, \mathrm{d}t. \tag{2.4}$$
2. As terminology indicates, $f^*$ is indeed radially-symmetric and decreases radially.
3. The "layer-cake" representation for $L^\alpha$-norms shows that if $f$ is a probability density, then so is $f^*$. Moreover, $h_\alpha(f) = h_\alpha(f^*)$ holds for all $\alpha > 1$.

For more information regarding rearrangements, we refer to Lieb and Loss' text [70, Chapter 3] and Burchard's notes [26].

### 2.2.1 Comparing Rényi entropies via majorization

Recall that we are trying to show $h_\alpha(T(X) + W) \leq h_\alpha(X + W)$ under various hypotheses. Towards this, we will try to show that $f^*_{T(X)+W}$ is less spread out than $f^*_{X+W}$. We formulate this notion of *spread* using the majorisation order.

**Definition 7.** For two probability densities $f$ and $g$ on $\mathbb{R}^d$, we say that $f$ is majorised by $g$, written as $f \preceq g$ (or $g \succeq f$), if

$$\int_{\mathcal{B}(0,r)} f^*(x) \, \mathrm{d}x \leq \int_{\mathcal{B}(0,r)} g^*(x) \, \mathrm{d}x,$$

for all $r > 0$. If $f \preceq g$ and $g \preceq f$, we will write $f \simeq g$.

Indeed, knowing $f \preceq g$ allows us to conclude that $h_\alpha(f) \geq h_\alpha(g)$. This can be seen by applying the lemma below to convex functions $\phi(x) = x^\alpha$, if $\alpha \geq 1$, and $\phi(x) = -x^\alpha$, if $\alpha \leq 1$.

**Lemma 2.2.1.** *[104, Lemma VII.2.] Let $\phi(x)$ be a convex function defined on the non-negative real line such that $\phi(0) = 0$ and is continuous at 0. If $f$ and $g$ are probability densities, with $f \preceq g$, then*

$$\int_{\mathbb{R}^d} \phi(f(x))\,\mathrm{d}x \leq \int_{\mathbb{R}^d} \phi(g(x))\,\mathrm{d}x.$$

To establish $f \preceq g$ for $f = f_{X+W}$ and $g = f_{T(X)+W}$, it is useful to have a more tractable representation of the integrals involved in Definition 7. The following elementary lemma is suitable for this purpose, which can also be found in [26].

**Lemma 2.2.2.** *Let $f : \mathbb{R}^d \to [0, \infty)$ be an integrable non-negative function. Then,*

$$\int_{B(0,r)} f^*(x)\,\mathrm{d}x = \sup_{\{C:\mathrm{Vol}(C)=\mathrm{Vol}(\mathcal{B}(0,r))\}} \int_C f(x)\,\mathrm{d}x. \tag{2.5}$$

*Moreover, the supremum is attained by any super-level set $\{f > t\}$ with the same volume as $\mathcal{B}(0, r)$.*

*Proof.* By the Hardy-Littlewood inequality (see, e.g., [70, Theorem 3.4]), one has

$$\int_C f(x)\,\mathrm{d}x \leq \int_{C^*} f^*(x)\,\mathrm{d}x.$$

Whence, we have,

$$\int_{B(0,r)} f^*(x)\,\mathrm{d}x \geq \sup_{\{C:\mathrm{Vol}(C)=\mathrm{Vol}(\mathcal{B}(0,r))\}} \int_C f(x)\,\mathrm{d}x.$$

To prove the reversed inequality, for any $r > 0$, by the property in Equation 2.4, there exists $t > 0$ (depending on $r$) such that $B(0, r) = \{x : f^*(x) > t\}$. Hence,

$$\int_{B(0,r)} f^*(x)\, \mathrm{d}x = \int_{\{x : f^*(x) > t\}} f^*(x)\, \mathrm{d}x = \int_0^\infty \mathrm{Vol}(\{x : f^*(x) > \max\{t, s\}\})\, \mathrm{d}s$$

$$= \int_0^\infty \mathrm{Vol}(\{x : f(x) > \max\{t, s\}\})\, \mathrm{d}s = \int_{\{x : f(x) > t\}} f(x)\, \mathrm{d}x.$$

The reversed inequality now follows. □

Using these pieces, along with an application of the Prékopa-Leindler inequality, we obtain the first result of this section.

**Theorem 2.2.3.** *For any two log-concave random vectors $X$ and $W$, and any $\lambda \in (0, 1)$, we have*

$$f_{X+W} \preceq f_{\lambda X+W},$$

*and consequently,*

$$h_\alpha(\lambda X + W) \leq h_\alpha(X + W) \text{ for all } \alpha \in (0, \infty).$$

*Proof.* Since $X$ and $W$ are log-concave, a direct application of the Prékopa-Leindler inequality reveals that $X + W$ is also log-concave. Consequently, the super level sets of $X + W$ are bounded and convex. By Lemma 2.2.2, it thus suffices to show that for any bounded convex $K$, there exists a Borel measurable set $K'$ of equal Lebesgue measure such that

$$\mathbb{P}\{X + W \in K\} \leq \mathbb{P}\{\lambda X + W \in K'\}.$$

By conditioning on $X$, it suffices to show that

$$\mathbb{P}\{x + W \in K\} \leq \mathbb{P}\{\lambda x + W \in K'\},$$

for arbitrary fixed $x$.

To that end, note that for a fixed bounded convex set $K$, an application of Prékopa-Leindler inequality shows that $p(x) := \mathbb{P}\{x+W \in K\}$ is a log-concave function of $x$. In particular, $p$ is unimodal. Denote by $x_0$ the point where the function $p$ achieves its maximum. By unimodality, we have, for any $\lambda \in [0, 1]$,

$$p\left((1-\lambda)x_0 + \lambda x\right) \geq p(x)$$

for an arbitrary fixed $x$.

In other words, we have the following for any arbitrary fixed $x$:

$$\mathbb{P}\{x + W \in K\} \leq \mathbb{P}\{(1-\lambda)x_0 + \lambda x + W \in K\}.$$

Setting $K' = K - (1-\lambda)x_0$, we thus obtain:

$$\mathbb{P}\{x + W \in K\} \leq \mathbb{P}\{\lambda x + W \in K'\},$$

as desired. □

Note that the geometric consequence of the previous result obtained by letting $X \sim \text{Uniform}(K)$, $W \sim \text{Uniform}(\mathcal{B})$ and $\alpha \to 0$, is trivial because $\lambda K + \mathcal{B} \subseteq K + \mathcal{B}$ up to a translation. The next result is similar in this aspect. But, as before, the entropic version needs a little more work.

For an arbitrary contraction $T$, if it acts on a ball $\mathcal{B}(0, r)$, then up to a shift, we would have $T(\mathcal{B}(0, r)) \subset \mathcal{B}(0, r)$. Therefore, if both $X$ and $W$ are radially-symmetric and unimodal, heuristically, one may expect that, up to a shift, the distribution of $T(X) + W$ is more concentrated than the distribution of $X + W$, hence should have smaller Rényi entropies. We turn this intuition into the following theorem:

**Theorem 2.2.4.** *For any radially-symmetric, unimodal, $\mathbb{R}^d$-valued random vectors $X$*

*and $W$, and for any contraction $T : \mathbb{R}^d \to \mathbb{R}^d$, we have*

$$f_{X+W} \preceq f_{T(X)+W},$$

*and consequently,*

$$h_\alpha(T(X) + W) \leq h_\alpha(X + W) \text{ for all } \alpha \in (0, \infty).$$

*Proof.* The density $f_{X+W}$ is already radially-symmetric and unimodal, so $f^*_{X+W} = f_{X+W}$. Given this, to show that $f_{X+W} \preceq f_{T(X)+W}$, by Lemma 2.2.2, we need to produce for each $r > 0$, a measurable set $B'$ with $\mathrm{Vol}(B') = \mathrm{Vol}(\mathcal{B}(0, r))$ satisfying $\mathbb{P}\{X + W \in \mathcal{B}(0, r)\} \leq \mathbb{P}\{T(X) + W \in B'\}$ . Anderson's theorem [4] implies that the function $x \mapsto \mathbb{P}\{x + W \in \mathcal{B}(0, r)\}$ is radially-symmetric and unimodal. Consequently, $\mathbb{P}\{x + W \in \mathcal{B}(0, r)\} \leq \mathbb{P}\{(T(x) - T(0)) + W \in \mathcal{B}(0, r)\}$ for each fixed $x$, since $T$ is a contraction. By conditioning,

$$\mathbb{P}\{X + W \in \mathcal{B}(0, r)\} \leq \mathbb{P}\{(T(X) - T(0)) + W \in \mathcal{B}(0, r)\},$$

and therefore, setting $B' = \mathcal{B}(0, r) + T(0)$ complete the proof.

□

*Remark.* Let $K$ be a compact set with non-zero volume, $K^*$ the centred ball with same volume as $K$, and $T$ a contraction as before. Now the Brunn-Minkowski inequality implies $\mathrm{Vol}(K + \mathcal{B}) \geq \mathrm{Vol}(K^* + \mathcal{B})$ in this case. Moreover, since $K^*$ is a ball, $T[K^*] - T(0) \subseteq K^*$ and hence $(T[K^*] - T(0)) + \mathcal{B} \subseteq K^* + \mathcal{B}$. Combining these two elementary observations we get, $\mathrm{Vol}(T[K^*] + \mathcal{B}) \leq \mathrm{Vol}(K + \mathcal{B})$. This inequality for volumes can also be obtained as a corollary to the above theorem by applying it to the case when $X \sim \mathrm{Uniform}(K), W \sim \mathrm{Uniform}(\mathcal{B})$, and using the observation due to Brascamp and Lieb that $h_\alpha(f \star g) \geq h_\alpha(f^* \star g^*)$ for $\alpha \in (0, 1)$ [21, Proposition 9].

It is possible to trade the condition on the radial symmetry of $X$ for certain (stronger) linearity assumptions on $T$. Recall that a random vector $X = (X_1, \dots . X_d)$ on $\mathbb{R}^d$ is said to be unconditional if all $(\pm X_1, \cdots, \pm X_d)$ have the same distribution regardless of the choice of signs $\pm$.

**Theorem 2.2.5.** *Let $X$ be an $\mathbb{R}^d$-valued log-concave random vector, $W$ an $\mathbb{R}^d$-valued unconditional log-concave random vector. For any diagonally linear contraction $T$, we have*

$$f_{X+W} \preceq f_{T(X)+W},$$

*and consequently,*

$$h_\alpha(T(X)+W) \leq h_\alpha(X+W) \textit{ for all } \alpha \in (0,\infty).$$

*Proof.* We will denote the density of $X$ and $W$ by $f$ and $g$ respectively. Since $W$ is unconditional, we may assume that the diagonal elements $\lambda_i$ of $T$ non-negative. Since both $X$ and $W$ are log-concave, $X + W$ is also log-concave and so the super-level sets $\{f_{X+W} > t\}$ of its density are convex sets. By appealing to the equality case in Lemma 2.2.2, it is sufficient to show that, for every convex set $K$, there exists a Borel measurable set $K'$ having the same volume as $K$, such that

$$\int_K f_{X+W} \leq \int_{K'} f_{T(X)+W}.$$

Upon explicitly writing out the convolution and changing variables, the above reads

$$\int_{K'} \int_{\mathbb{R}^d} f(y) g(x - T(y)) \, \mathrm{d}y \, \mathrm{d}x \geq \int_K \int_{\mathbb{R}^d} f(y) g(x-y) \, \mathrm{d}y \, \mathrm{d}x,$$

where $K'$ is a Borel measurable set having the same Lebesgue measure as $K$.

To this end, by Fubini's theorem, we have

$$\int_K g(x - T(y))\ \mathrm{d}x =$$
$$\int_{\Pi_{e_1^\perp}(K)} \left( \int_{I(x_2,\cdots,x_d)} g(x_1 - \lambda_1 y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x_1 \right)\ \mathrm{d}x_2 \cdots\ \mathrm{d}x_d,$$

where $\Pi_{e_1^\perp}(K)$ denotes orthogonal projection of $K$ onto the orthogonal complement of $e_1$ and $I(x_2, \cdots, x_d)$ is the support of the inner integrand. Note that, since $K$ is convex, for fixed $x_2, \cdots, x_d$ and $y$,

$$I(x_2, \cdots, x_d) = [a(x_2, \cdots, x_d), b(x_2, \cdots, x_d)]$$

is an interval, and $g(z, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)$ is an even log-concave function in $z$. Hence,

$$p(z) := \int_{I(x_2,\cdots,x_d)} g(x_1 - z, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x_1,$$

is a log-concave function on $\mathbb{R}$ whose maximum is attained at

$$z_0 = \frac{a(x_2, \cdots, x_d) + b(x_2, \cdots, x_d)}{2}.$$

We deduce that

$$p((1 - \lambda_1) z_0 + \lambda_1 y_1) \geq p(y_1).$$

Whence we have

$$\int_{I(x_2,\cdots,x_d)-(1-\lambda_1)z_0} g(x_1 - \lambda_1 y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x_1$$

$$\geq \int_{I(x_2,\cdots,x_d)} g(x_1 - y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x_1.$$

To summarize, we have shown that

$$\int_{S_{e_1}^{\lambda_1}(K)} g(x_1 - \lambda_1 y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x \geq$$

$$\int_K g(x_1 - y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d)\ \mathrm{d}x,$$

where $S_{e_i}^{\lambda_i}(K)$ are defined as the following:

$$\left\{ x \times \left\{ \left[ -\frac{t_2 - t_1}{2}, \frac{t_2 - t_1}{2} \right] + \lambda_i \frac{(t_2 + t_1)}{2} \right\} : x \in e_i^{\perp}, (x, te_i) \cap K = x \times [t_1, t_2] \right\}.$$

It is worth noting that, $S_{e_i}^{\lambda_i}(K)$ is a member of the so-called shadow system of $K$ along the direction $e_i$, which was introduced in [93]. It is well-known that shadow systems preserve convexity, however we provide a proof here for completeness. Let $u$ be any unit vector, note that the convex body $K$ can be expressed as

$$K = \{(x, su) : x \in u^{\perp},\ g(x) \leq s \leq f(x)\},$$

where $g(x)$ is a convex function and $f(x)$ is a concave function. Therefore, $S_u^{\lambda}(K)$ is

$$\left\{ x \times \left\{ \left[ -\frac{f(x) - g(x)}{2}, \frac{f(x) - g(x)}{2} \right] + \lambda \frac{g(x) + f(x)}{2} \right\} : x \in u^{\perp}, \lambda \in [0, 1] \right\}.$$

Since $\lambda \in [0, 1]$, we have that $-\frac{f(x)-g(x)}{2} + \lambda \frac{g(x)+f(x)}{2}$ is concave, while $\frac{f(x)-g(x)}{2} + \lambda \frac{g(x)+f(x)}{2}$ is convex. This establishes the convexity of $S_u^{\lambda}(K)$. Moreover, by Fubini's theorem, $S_{e_i}^{\lambda_i}$ also preserves the volume:

$$\mathrm{Vol}(S_{e_i}^{\lambda_i}(K)) = \mathrm{Vol}(K), \quad i = 1, \cdots, d.$$

Repeating the argument coordinatewise, we have

$$\int_{S_{e_1}^{\lambda_1} S_{e_2}^{\lambda_2} \cdots S_{e_d}^{\lambda_d}(K)} g(x_1 - \lambda_1 y_1, x_2 - \lambda_2 y_2, \cdots, x_d - \lambda_d y_d) \, \mathrm{d}x \geq \int_K g(x_1 - y_1, x_2 - y_2, \cdots, x_d - y_d) \, \mathrm{d}x.$$

Choose $K' = S_{e_1}^{\lambda_1} S_{e_2}^{\lambda_2} \cdots S_{e_d}^{\lambda_d}(K)$, the desired result now follows. □

If we go back and check our proof for Theorem 2.2.5 , we essentially prove the following pointwise inequality.

**Theorem 2.2.6.** *For any convex body $K$ in $\mathbb{R}^d$, any diagonal matrix $T$ with diagonal entries in $[0,1]$, and $g(x)$ unconditional log-concave or radially symmetric unimodal, there exists a convex body $S(K)$ having same volume as $K$, such that*

$$\int_{S(K)} g(x + T(y)) \, \mathrm{d}x \geq \int_K g(x + y) \, \mathrm{d}x$$

*for all $y \in \mathbb{R}^d$.*

It is worth comparing with the following theorem due to Anderson [4].

**Theorem 2.2.7.** *Let $K$ be an origin symmetric convex body in $\mathbb{R}^d$ and let $g(x)$ be a nonnegative, symmetric, unimodal,and integral function on $\mathbb{R}^d$. Then*

$$\int_K g(x + cy) \, \mathrm{d}x \geq \int_K g(x + y) \, \mathrm{d}x,$$

*for all $0 \leq c \leq 1$ and $y \in \mathbb{R}^d$.*

Our theorem relaxes the condition for $K$ but impose stronger condition on $g$, and it essentially says that one may get an Anderson-type inequality by scaling the coordinates of $y$ at different rates, but has to pay a cost by rearranging the convex set $K$ accordingly.

If in addition $W$ is radially symmetric, then rotational invariance enables us to generalize Theorem 2.2.5 to any affine contraction $T$.

**Corollary 2.2.8.** *If $X$ is a log-concave random vector and $W$ is a radially-symmetric log-concave random vector, then for any affine contraction $T$, we have*

$$f_{X+W} \preceq f_{T(X)+W},$$

*and consequently,*

$$h_\alpha(T(X)+W) \leq h_\alpha(X+W) \text{ for all } \alpha \in (0,\infty).$$

*Proof.* First, note that the majorisation order remains invariant under orthogonal transformations (see Definition 7 and Equation (2.4)). By polar factorization and further diagonalisation of the symmetric component, we can write $T = Q_1 \Lambda Q_2$ for orthogonal matrices $Q_1, Q_2$ and diagonal $\Lambda$. Using Theorem 2.2.5,

$$\begin{aligned} f_{T(X)+W} = f_{Q_1\Lambda Q_2(X)+W} = f_{Q_1\Lambda Q_2(X)+Q_1 W} \simeq f_{\Lambda Q_2(X)+W} \succeq f_{Q_2(X)+W} \\ = f_{Q_2(X)+Q_2 W} \simeq f_{X+W}. \end{aligned}$$

□

By letting $\alpha \to 0$, we obtain the following inequality for convex bodies.

**Corollary 2.2.9.** *Let $K$ be a convex body and $r > 0$. Then,*

$$\mathrm{Vol}(T(K)+r\mathcal{B}) \leq \mathrm{Vol}(K+r\mathcal{B}),$$

*for any affine contraction $T$.*

This corollary can also be seen as a consequence of the fact that intrinsic volumes decrease under linear contractions [86, Proposition 1.1]. Interestingly, the technique that we use to prove Theorem 2.2.5 can be directly used to prove the Corollary 2.2.9 and the fact that intrinsic volumes decrease under linear contractions. We will revisit these results in the next setion.

Further, if we impose the stronger restriction of unconditionality on $X$, then Corollary 2.2.8 also holds when $T$ is a strong contraction.

**Theorem 2.2.10.** *Let $X, W$ be two unconditional log-concave random vectors. Then for any strong contraction $T$, we have*

$$f_{X+W} \preceq f_{T(X)+W},$$

*and consequently,*

$$h_\alpha(T(X)+W) \leq h_\alpha(X+W) \text{ for all } \alpha \in (0,\infty).$$

*Proof.* Denote by $f$ and $g$ the densities of $X$ and $W$, respectively. Note that Rényi entropy is translation invariant, by subtracting $T(0)$ from $T$, we may assume that $T(0) = 0$. Since both $X$ and $W$ are unconditional log-concave, $r(x)$, the density of $X + W$, is also unconditional log-concave, whose super level sets, therefore, are unconditional convex sets. Again by Lemma 2.2.1 and Lemma 2.2.2, it suffices to show that for any unconditional convex set $K$, one has

$$\int_K \int_{\mathbb{R}^d} f(y)g(x-T(y)) \text{ d}y \text{ d}x \geq \int_K \int_{\mathbb{R}^d} f(y)g(x-y) \text{ d}y \text{ d}x.$$

It suffices to show that

$$\int_K g(x-T(y)) \text{ d}x \geq \int_K g(x-y) \text{ d}x.$$

To this end, by Fubini's theorem, we have

$$\int_K g(x-T(y)) \text{ d}x =$$
$$\int_{\Pi_{e_1^\perp}(K)} \left( \int_{I(x_2,\cdots,x_d)} g(x_1 - T_1(y), x_2 - T_2(y), \cdots, x_d - T_d(y)) \text{ d}x_1 \right) \text{ d}x_2 \cdots \text{ d}x_d,$$

where $T$ is represented as $(T_1, \ldots, T_d)$. Note that, for fixed $x_2, \cdots, x_d$ and $y$, $I(x_2, \cdots, x_d)$ is a symmetric interval and $g(z, x_2 - T_2(y), \cdots, x_d - T_d(y))$ is an even log-concave function in $z$. Hence,

$$p(z) := \int_{I(x_2,\cdots,x_d)} g(x_1 - z, x_2 - T_2(y), \cdots, x_d - T_d(y)) \, \mathrm{d}x_1,$$

is an even log-concave function on $\mathbb{R}$. Now by the strong contractivity of $T$, we deduce that

$$p(T_1(y)) = p(T_1(y) - T_1(0)) \geq p(y_1).$$

Whence we have

$$\int_{I(x_2,\cdots,x_d)} g(x_1 - T_1(y), x_2 - T_2(y) \cdots, x_d - T_d(y)) \, \mathrm{d}x_1$$

$$\geq \int_{I(x_2,\cdots,x_d)} g(x_1 - y_1, x_2 - T_2(y) \cdots, x_d - T_d(y)) \, \mathrm{d}x_1.$$

To summarize, we have shown that

$$\int_K g(x_1 - T_1(y), x_2 - T_2(y), \cdots, x_d - T_d(y)) \, \mathrm{d}x \geq \int_K g(x_1 - y_1, x_2 - T_2(y), \cdots, x_d - T_d(y)) \, \mathrm{d}x.$$

Repeat the argument coordinatewise, we have

$$\int_K g(x_1 - T_1(y), x_2 - T_2(y), \cdots, x_d - T_d(y)) \, \mathrm{d}x \geq \int_K g(x_1 - y_1, x_2 - y_2, \cdots, x_d - y_d) \, \mathrm{d}x$$

as desired. □

**Corollary 2.2.11.** *Let $K, L$ be two unconditional convex bodies, then*

$$\mathrm{Vol}(T(K) + L) \leq \mathrm{Vol}(K + L)$$

*holds for any strong contraction $T$.*

Maps of the form $\nabla\varphi$, for convex $\varphi$, play an important role in geometry via the theory of optimal transport where such maps solve the mass transport problem for the quadratic cost (see for example, [102, Chapter 2]). Moreover, they play the role of the positive-semidefinite matrices in a far reaching generalization of polar factorisation of matrices to maps $\mathbb{R}^d \to \mathbb{R}^d$ due to Brenier [23]. We think it is worthwhile to note that a result for contractions of this form can be obtained if we assume $W$ to be radially-symmetric and log-concave.

**Corollary 2.2.12.** *If $X$ is unconditionally log-concave, $W$ is radially-symmetric log concave, and $T = \nabla\varphi$ for some smooth convex function $\varphi$ on $\mathbb{R}^d$, is a contraction. Then, we have*

$$f_{X+W} \preceq f_{T(X)+W},$$

*and consequently,*

$$h_\alpha(T(X)+W) \le h_\alpha(X+W) \text{ for all } \alpha \in (0,\infty).$$

*Proof.* Replacing $\varphi(x)$ with $\varphi(x) - \langle x, \nabla\varphi(0)\rangle$, if necessary, we may assume that $\nabla\varphi(0) = 0$. Since $\varphi$ is a smooth convex function, apply the multi-dimensional mean value theorem, we have

$$\nabla\varphi(y) - \nabla\varphi(0) = \left(\int_0^1 DT(ty)\ \mathrm{d}t\right)\cdot y.$$

For each $y \in \mathbb{R}^d$,

$$H(y) := \int_0^1 DT(ty)\ \mathrm{d}t,$$

as a convex combination of positive semi-definite matrices, is again positive semi-definite. Therefore, there exists an orthogonal matrix $Q_y$, such that $H(y) = Q_y^\perp \Lambda_y Q_y$, where $\Lambda_y$ is a diagonal matrix for every $y$. Then for any symmetric convex set $K$,

$$\int_K g(x - \nabla\varphi(y))\ \mathrm{d}x = \int_K g(x - H(y)\cdot y)\ \mathrm{d}x = \int_K g(x - (Q_y^{\perp}\Lambda_y Q_y)\cdot y)\ \mathrm{d}x.$$

By radial-symmetry of $g$ and Theorem 2.2.10 applied to $\Lambda_y$, we have

$$\int_K g(x - (Q_y^{\perp}\Lambda_y Q_y)y)\ \mathrm{d}x = \int_{Q_y(K)} g(x - \Lambda_y\cdot((Q_y)y))\ \mathrm{d}x.$$

$$\geq \int_{Q_y(K)} g(x - (Q_y)\cdot y)\ \mathrm{d}x = \int_K g((Q_y)\cdot x - (Q_y)\cdot y)\ \mathrm{d}x = \int_K g(x-y)\ \mathrm{d}x,$$

as desired.

□

### 2.2.2 Mixed volumes and diagonal linear contractions

#### 2.2.2.1 Mixed volumes

The definition of mixed volumes was first introduced by Minkowski in [84]. In this subsection, we quote the main theorem of Minkowski on the polynomiality of volume with respect to Minkowski addition.

**Theorem 2.2.13.** *Let $K_1,\dots,K_m$ be non-empty compact convex subsets of $\mathbb{R}^n$. For any $n$-tuple $1\leq i_1,\dots,i_n\leq m$ there exists non-negative coefficients $V(K_{i_1},\dots,K_{i_n})$, $1\leq i_1,\dots,i_n\leq m$, that are symmetric with respect to the indices $i_1,\dots,i_n$, such that*

$$\mathrm{Vol}_n(t_1K_1+\ldots+t_mK_m) = \sum_{i_1,\dots,i_n=1}^{m} V(K_{i_1},\dots,K_{i_n})t_{i_1}\dots t_{i_n}$$

*for all $t_1\dots t_m\geq 0$. The coefficient $V(K_{i_1},\dots,K_{i_n})$ is called the mixed volume of $K_{i_1},\dots,K_{i_n}$ and depends only on these $n$ bodies. We use $V(K,[m],L,[n-m])$ to denote the mixed volume of $m$ copies of $K$ with $n-m$ copies of $L$.*

A special case of the above is the formula for $\mathrm{Vol}_n(K + t\mathcal{B})$, for $t > 0$, can be expanded as a polynomial in $t$:

$$\mathrm{Vol}_n(K + t\mathcal{B}) = \sum_{i=0}^{n} \binom{n}{i} W_i(K) t^i = \sum_{i=0}^{n} k_{n-j} V_i(K) t^{n-i},$$

where $W_i(K) = V(K, [n-i], \mathcal{B}, [i])$ is called the $i$-th quermassintegral of $K$, $V_i(K)$ is called the $i$-th intrinsic volume of $K$, and $k_m$ denotes the volume of the Euclidean unit ball in $\mathbb{R}^m$.

We list a few properties of mixed volumes, which can be found in [6]. Below, $K$,$L$, $K_i$ are convex bodies.

(i) $V(K_1, \ldots, K_n) = V(K_{\sigma(1)}, \ldots, K_{\sigma(n)})$ for any permutation $\sigma$.

(ii) $V$ is multilinear, that is, linear in each argument: for $\lambda, \mu > 0$,

$$V(\lambda K + \mu L, K_2 \ldots, K_n) = \lambda V(K, K_2 \ldots, K_n) + \mu V(L, K_2 \ldots, K_n)$$

(iii) $V(U(K_1), U(K_2) \ldots, U(K_n)) = V(K_1, K_2 \ldots, K_n)$ for $U \in O(n)$.

#### 2.2.2.2 Shadow systems

Let $a$ be a unit vector in $\mathbb{R}^{n+1}$, and denote by $a^\perp$ the $n$ dimensional subspace that is perpendicular to $a$, also let $K$ be a convex body in $\mathbb{R}^{n+1}$. Then the system of convex bodies in $\mathbb{R}^n$ defined by

$$K^a(u) := \Pi_{a+u}(K), \;\; u \in a^\perp,$$

where $\Pi_{a+u} : \mathbb{R}^{n+1} \to a^\perp$ denotes the projection along the line spanned by the vector $a + u$. To be more precise, $\Pi_{a+u}(K) = \{x + \lambda(a+u) : x \in K, \lambda \in \mathbb{R}\} \cap a^\perp$. $K(u)$ is called the shadow system induced by $K$, which was first introduced by Shephard [96]. It is worth noting that if $K$ is a convex body in $a^\perp$, then one has $K^a(u) = K$ for $u \in a^\perp$. The following theorem plays a key role in our proof.

**Theorem 2.2.14.** *[96] If $u \mapsto K_1^a(u), \ldots, K_n^a(u)$ are shadow systems, then the function $u \mapsto V(K_1^a(u), \ldots, K_n^a(u))$ is convex.*

To see how the steiner symmetrization is connected to a shadow system, we follow the exposition in [94]. One first represents a compact convex body $K$ in $\mathbb{R}^n$ in the following way

$$K = \{y + \alpha u : y \in K|_{u^\perp},\ g(y) \le \alpha \le f(y)\},$$

where $u$ is an arbitrary unit vector and $K|_{u^\perp}$ denotes the projection of $K$ onto the hyperplane $u^\perp$.

Then for this fixed unit vector $u$, one defines a convex set in $\mathbb{R}^{n+1}$ as follows:

$$\widetilde{K} := \overline{\text{conv}}\{(x, \beta(x|_{u^\perp})) : x \in K\},$$

where $\beta(y) = -[g(y) + f(y)]$ for $y \in K|_{u^\perp}$.

Let $a = (0, \ldots, 0, -1)$ be a vector in $\mathbb{R}^{n+1}$, define a shadow system in the following way:

$$K(tu) := \widetilde{K}^a(t(u, 0)) = \overline{\text{conv}}\{x + t\beta(x|_{u^\perp})u : x \in K\}.$$

By convexity of $K$, one can deduce that $g(y)$ is convex and $f(y)$ is concave on $K|_{u^\perp}$. Therefore, for $t \in [0, 1]$, we have that $f(y) - t[g(y) + f(y)]$ is concave, and $g(y) - t[g(y) + f(y)]$ is convex. In other words, for $t \in [0, 1]$, $\{x + t\beta(x|_{u^\perp})u : x \in K\}$ is convex. Hence,

$$K(tu) = \{x + t\beta(x|_{u^\perp})u : x \in K\}.$$

Denote by $\sigma_{u^\perp}$ the reflection map with repect to the $n-1$ dimensional subspace $u^\perp$. Note that $K(0) = K$, $K(u) = \sigma_{u^\perp}(K)$ and $K(u/2)$ gives us the usual steiner symmetrization. Moreover, by Fubini's theorem, one has $\text{Vol}(K(tu)) = \text{Vol}(K)$ for all $t \in [0, 1]$. The following inclusion relation is often found to be useful.

Let $K_1$ and $K_2$ be two convex bodies, then we have, for $t \in [0,1]$,

$$K_1(tu) + K_2(tu) \subseteq (K_1 + K_2)(tu). \tag{2.6}$$

Indeed, one can represent the convex bodies $K_i$ as follows:

$$K_i = \{y + \alpha u : y \in K_i|_{u^\perp},\ g_i(y) \le \alpha \le f_i(y)\},\ i = 1, 2.$$

Then by definition

$$K_i(tu) = \{y + \alpha u : y \in K|_{u^\perp},\ \alpha \in \{(1-t)[g_i(y), f_i(y)] + t[-f_i(y), -g_i(y)]\}\},\ i = 1, 2. \tag{2.7}$$

Similarly, $K_1 + K_2$ can be represented as follows:

$$K_1 + K_2 = \{y + \alpha u : y \in (K_1 + K_2)|_{u^\perp},\ g_3(y) \le \alpha \le f_3(y)\}.$$

And

$$(K_1+K_2)(tu) = \{y + \alpha u : y \in (K_1 + K_2)|_{u^\perp},\ \alpha \in \{(1-t)[g_3(y), f_3(y)] + t[-f_3(y), -g_3(y)]\}\} \tag{2.8}$$

Note that for $y_i \in K_i|_{u^\perp}$, one has that both $y_1 + y_2 + [g_1(y_1) + g_2(y_2)]u$ and $y_1 + y_2 + [f_1(y_1) + f_2(y_2)]u$ belong to $K_1 + K_2$. Therefore, we deduce that

$$g_1(y_1) + g_2(y_2) \ge g_3(y_1 + y_2),\ f_1(y_1) + f_2(y_2) \le f_3(y_1 + y_2). \tag{2.9}$$

Combine (2.7), (2.8), and inequalities (2.9), we deduce that

$$K_1(tu) + K_2(tu) \subseteq (K_1 + K_2)(tu),$$

as desired.

#### 2.2.2.3 Mixed volumes decrease under diagonal linear contractions

**Theorem 2.2.15.** *Let $K_i$ be unconditional convex bodies in $\mathbb{R}^n$, and let $\Lambda$ be a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n \in [-1,1]$. Then, for any convex body $K$ in $\mathbb{R}^n$, one has*

$$V(\Lambda(K), [n-j], K_1, \dots, K_j) \leq V(K, [n-j], K_1, \dots, K_j), \; j = 0, \dots, n.$$

*Proof.* Let $\Lambda_1$ be the diagonal matrix with diagonal entries being $\lambda_1, 1, \dots, 1$, and denote by $e_j$ the unit vector in $\mathbb{R}^n$ with $j$-th entry being 1. For $t \in [0,1]$, consider the shadow system

$$K(te_1) = \{x + t\beta(x|_{e_1^\perp})e_1 : x \in K\}.$$

For $t_{\lambda_1} = \frac{1-\lambda_1}{2}$, one has that

$$K(t_{\lambda_1}e_1) = \left\{ y + \alpha e_1 : y \in K|_{e_1^\perp}, \; \frac{1+\lambda_1}{2} g(y) - \frac{1-\lambda_1}{2} f(y) \leq \alpha \leq \frac{1+\lambda_1}{2} f(y) - \frac{1-\lambda_1}{2} g(y) \right\}$$

$$= \left\{ y + \alpha e_1 : y \in K|_{e_1^\perp}, \; \alpha \in \left\{ \frac{1+\lambda_1}{2}[g(y), f(y)] + \frac{1-\lambda_1}{2}[-f(y), -g(y)] \right\} \right\}$$

$$\supseteq \left\{ y + \alpha e_1 : y \in K|_{e_1^\perp}, \; \alpha \in \lambda_1[g(y), f(y)] \right\} = \Lambda_1(K).$$

Since $K_i$ are unconditional, one has

$$V(K, [n-j], K_1, \dots, K_j) = V(\sigma_{e_1^\perp}(K), [n-j], K_1, \dots, K_j).$$

Therefore, we deduce that

$$V(K(0), [n-j], K_1, \dots, K_j) = V(K(e_1), [n-j], K_1, \dots, K_j), \; j = 0, \dots, n. \quad (2.10)$$

By Theorem 2.2.14, one has that $V(K(te_1), [n-j], K_1, \dots, K_j)$ is a convex function in

$t$ for $t \in [0,1]$. Therefore, equation (2.10) implies that, for $\lambda_1 \in [-1,1]$,

$$V(K,[n-j],K_1,\ldots,K_j) \geq V(K(t_{\lambda_1}e_1),[n-j],K_1,\ldots,K_j).$$

By monotonicity of mixed volumes, we have

$$V(K(t_{\lambda_1}e_1),[n-j],K_1,\ldots,K_j) \geq V(\Lambda_1(K),[n-j],K_1,\ldots,K_j).$$

Therefore,

$$V(K,[n-j],K_1,\ldots,K_j) \geq V(\Lambda_1(K),[n-j],K_1,\ldots,K_j).$$

Once noticing that $\Lambda = \Lambda_n \cdots \Lambda_1$, one may repeat the argument consecutively to complete the proof. □

The proof above shows that for any convex body $K$, and any diagonal linear contraction $\Lambda$, there exists a sequence of (asymmetric) steiner symmetrizations $S_{e_i}^{\lambda_i}$,

$$S_{e_i}^{\lambda_i}(x) := x|_{e_i^\perp} + \left( x_i - \frac{1-\lambda_i}{2}[g(x|_{e_i^\perp}) + f(x|_{e_i^\perp})] \right) e_i,\ x = (x_1,\ldots,x_n) \in K,$$

such that

$$\Lambda(K) \subseteq \underbrace{S_{e_n}^{\lambda_n} \ldots S_{e_1}^{\lambda_1}}_{S}(K). \tag{2.11}$$

Note that for any Euclidean ball $\mathcal{B}$ centered at the origin, one has $S(\mathcal{B}) = \mathcal{B}$. And by the inclusion relation (2.6), one has that, for any convex body $K$ in $\mathbb{R}^n$,

$$S(K) + S(\mathcal{B}) \subseteq S(K+\mathcal{B}). \tag{2.12}$$

We immediately get the following corollary.

**Corollary 2.2.16.** *Let $K$ be any convex body in $\mathbb{R}^n$, $T$ be any linear contraction, and*

*$\mathcal{B}$ be the unit ball in $\mathbb{R}^n$ centered at the origin. Then*

$$\mathrm{Vol}(T(K) + \mathcal{B}) \leq \mathrm{Vol}(K + \mathcal{B}).$$

*Proof.* By singular value decomposition, there exist matrices $P, Q \in O(n)$ such that $T = P\Lambda Q$, where $\Lambda$ is a diagonal matrix with diagonal entries belong to the interval $[0, 1]$. Using the rotational invariance of Lebesgue measure, inculsion relation 2.11 and 2.12, and the volume preserving property of (asymmetric) steiner symmetrization, we obtain

$$\mathrm{Vol}(T(K) + \mathcal{B}) = \mathrm{Vol}(\Lambda(K) + \mathcal{B}) \leq \mathrm{Vol}(S(K) + \mathcal{B})$$

$$= \mathrm{Vol}(S(K) + S(\mathcal{B})) \leq \mathrm{Vol}(S(K + \mathcal{B})) = \mathrm{Vol}(K + \mathcal{B}).$$

□

The following corollary is an immediate consequence of Theorem 2.2.15.

**Corollary 2.2.17.** *Let $K$ be any convex body in $\mathbb{R}^n$ and $T$ be any linear contraction. Then*

$$V_i(T(K)) \leq V_i(K), \; i = 0, \ldots, n.$$

*Proof.* By definition of $V_i$, one has

$$V_i(K) = c_{n,i} V(K, [i], \mathcal{B}, [n - i]),$$

where $c_{n,i}$ are explicit constants (see, e.g., [94]).

Let $T = P\Lambda Q$ be its singular value decomposition. To prove the desired inequality, by the rotational invariance of $\mathcal{B}$, it suffices to show that

$$V(\Lambda(K), [i], \mathcal{B}, [n - i]) \leq V(K, [i], \mathcal{B}, [n - i]),$$

which follows from Theorem 2.2.15 immediately. □

## 2.3 Gaussian noise

The entropy power inequality (EPI),

$$N(X+Y) \geq N(X) + N(Y), \tag{2.13}$$

for $\mathbb{R}^d$-valued random vectors $X, Y$ with density, is a fundamental result of information theory occupying a place akin to the Brunn-Minkowski inequality in convex geometry. Here $N(X) := e^{\frac{2h(X)}{d}}$ denotes the entropy power of $X$. It was shown by Costa [28] that this inequality improves when one of the random variables involved is Gaussian. More precisely, he observed that $N(X + \sqrt{t}Z)$ is concave as a function of $t$. Later, a vector-generalization[1] of Costa's inequality was published by Liu, Liu, Poor and Shamai [71]. However, a flaw in their proof and a partial resolution was discovered by Courtade, Han and Wu [30]. Thankfully, we will only need the "correct part" of the vector-generalization.

**Theorem 2.3.1.** *[30, 71] Let $Z \sim \mathcal{N}(0, \Sigma)$ be a Gaussian random vector in $\mathbb{R}^d$ and $X$ be any random vector in $\mathbb{R}^d$ having density with respect to the Lebesgue measure. Suppose $S$ is a positive-semidefinite matrix such that $I_d - S$ is also positive-semidefinite and $S$ commutes with $\Sigma$. Then,*

$$N(X + S^{1/2}Z) \geq \det(I-S)^{1/d} N(X) + \det S^{1/d} N(X+Z).$$

Using these results, we can solve the Gaussian version of the Kneser-Poulsen problem for the Boltzmann-Shannon entropy in a strong sense when the contraction $T$ is linear.

**Theorem 2.3.2.** *Let $Z \sim \mathcal{N}(0, I_d)$ be the standard Gaussian random vector in $\mathbb{R}^d$, $X$ a random vector in $\mathbb{R}^d$ having density with respect to the Lebesgue measure, and $T$ a*

[1] here "vector-generalization" refers to the usage of a matrix instead of the "$t$" in Costa's EPI

*linear contraction. Then we have*

$$N(X+Z) \geq N(T(X)+Z) + (1-\mathrm{Lip}^2(T))N(X),$$

*where* $\mathrm{Lip}(T)$ *denotes the Lipschitz constant of the contraction* $T$*.*

*Proof.* By polar decomposition, there exists an orthogonal matrix $Q$ and a positive semi-definite matrix $A$ such that $T = QA$. Following the invariance of the entropy power and the invariance of the standard Gaussian distribution under the action of the orthogonal group, we have $N(T(X)+Z) = N(A(X)+Z)$. Moreover, since $T$ is contractive, $I - A^2$ is positive-semidefinite. Therefore, by the vector version of the EPI (2.3.1) applied to $S = A^2$ and $A(X)$ instead of $X$,

$$N(A(X)+A(Z)) \geq \det(I_d - A^2)^{1/d} N(A(X)) + \det(A^2)^{1/d} N(A(X)+Z).$$

Assume first that $T$ is non-singular so that $A$ is positive-definite. Then an application of the change of variables formula $h(A(X)) = h(X) + \overline{E} \log\det A'(X)$ yields

$$N(X+Z) \geq \det(I_d - A^2)^{1/d} N(X) + N(A(X)+Z).$$

Denote by $\lambda_1(A)$ the largest eigenvalue of $A$, we have that all eigenvalues of $I_d - A^2$ are greater than $1 - \lambda_1^2(A)$. The desired inequality now follows from the fact that $\lambda_1(A)$ is also equal to the Lipschitz constant of the linear map $T$.

If $T$ is singular, by the argument above, we may assume that $T$ is positive semi-definite. Therefore, there exists a sequence of positive definite matrices $T_\epsilon$ such that $T_\epsilon \to T$, as $\epsilon \to 0$. For such $T_\epsilon$,

$$N(X+Z) \geq N(T_\epsilon(X)+Z) + (1-\sigma_1^2(T_\epsilon))N(X).$$

Since $\sigma_1^2(T_\epsilon) \to \sigma_1^2(T)$, as $\epsilon \to 0$, all we need to show is that

$$N(T_\epsilon(X) + Z) \to N(T(X) + Z), \quad \text{as } \epsilon \to 0.$$

Denote by $f(x)$ and $f_\epsilon(x)$ the densities of $T(X) + Z$ and $T_\epsilon(X) + Z$, respectively. By the Dominated Convergence Theorem, one has that

$$f_\epsilon(x) \to 0 \quad \text{uniformly as } \|x\| \to \infty,$$

$$f_\epsilon(x) \quad \text{are uniformly bounded},$$

$$f_\epsilon(x) \to f(x) \quad \text{uniformly as } \epsilon \to 0.$$

We may assume that $\sup_\epsilon h(T_\epsilon(X) + Z) < \infty$ . Otherwise, by the previous result, we would have $N(X + Z) = \infty$, nothing needs to be proved. Then by Fatou's Lemma, we have $h(T(X) + Z) < \infty$. Note that we may also assume that $h(T(X) + Z) > -\infty$, otherwise, the desired inequality holds trivially. To summarize, we want to show that

$$h(T_\epsilon(X) + Z) \to h(T(X) + Z), \quad \text{as } \epsilon \to 0,$$

under the assumptions that $h(T(X) + Z) \in \mathbb{R}$ and $\sup_\epsilon h(T_\epsilon(X) + Z) < \infty$, which now can be verified readily. □

**Theorem 2.3.3.** *Let $Z \sim \mathcal{N}(0, I_d)$ be the standard Gaussian random vector in $\mathbb{R}^d$, $G \sim \mathcal{N}(\mu, \Lambda)$ be any Gaussian random vector in $\mathbb{R}^d$ with diagonal covariance matrix and $T$ any strong contraction. Then we have*

$$N(G + Z) \geq N(T(G) + Z) + (1 - \mathrm{Lip}^2(T))N(G).$$

*If, in addition, $\Lambda = \alpha I_d$ for some $\alpha > 0$, the inequality holds for any contraction $T$.*

*Proof.* A direct calculation reveals that

$$N(G+Z) = 2\pi e \det(I_d + \Lambda)^{1/d}, \quad N(G) = 2\pi e \det(\Lambda)^{1/d}.$$

Since $(1-\mathrm{Lip}^2(T))\Lambda$ is positive semi-definite, by the fact that $\det(A)^{1/d}$ is concave on the set of positive semi-definite matrix, one has

$$\det(I_d+\Lambda)^{1/d} \geq \det(I_d + \mathrm{Lip}^2(T)\Lambda)^{1/d} + \det((1-\mathrm{Lip}^2(T))\Lambda)^{1/d}.$$

Denote by $\Sigma$ the covariance matrix of $T(G)$, by the well-known fact that Gaussians maximize entropy under second-moment constraints, one has

$$N(T(G)+Z) \leq 2\pi e \det(I_d + \Sigma)^{1/d}.$$

Therefore, to prove the desired inequality, it suffices to show that

$$\det(I_d+\Sigma) \leq \det(I_d + \mathrm{Lip}^2(T)\Lambda).$$

Let $\lambda_1, \cdots, \lambda_d$ be the eigenvalues of $\Sigma$, and $a_{11}, \cdots, a_{nn}$ be the diagonal entries of $\Sigma$. The first observation is that

$$a_{ii} \leq \mathrm{Lip}^2(T)\Lambda_{ii}.$$

To see this, note that

$$a_{ii} = \mathbb{E}[T_i(G) - \mathbb{E}(T_i(G))]^2 \leq \mathbb{E}[T_i(G) - T_i(\mathbb{E}(G))]^2 \leq \mathrm{Lip}^2(T)\Lambda_{ii},$$

where the first inequality follows from the property of variance, while the second follows from the strong contractivity of $T$.

On the other hand,

$$\det(I_d+\Sigma) = \prod_{i=1}^{d}(1+\lambda_i) \leq \prod_{i=1}^{d}(1+a_{ii}) \leq \prod_{i=1}^{d}(1+\mathrm{Lip}^2(T)\Lambda_{ii}) = \det(I_d + \mathrm{Lip}^2(T)\Lambda),$$

where the first inequality is Hadamard's inequality for positive semi-definite matrices. If, in addition, $\Lambda = \alpha I_d$, by the AM-GM inequality, one has

$$\det(I_d + \Sigma)^{1/d} = \prod_{i=1}^{d}(1+\lambda_i)^{1/d} \leq 1 + \frac{\mathrm{Tr}(\Sigma)}{d}.$$

By the property of variance, we have

$$\begin{aligned}\mathrm{Tr}(\Sigma) = \mathbb{E}\|T(G) - \mathbb{E}(T(G))\|_2^2 \leq \\ \mathbb{E}\|T(G) - T(\mathbb{E}(G))\|_2^2 \leq \mathrm{Lip}^2(T)\mathrm{Tr}(\Lambda) = \mathrm{Lip}^2(T) d\alpha.\end{aligned}$$

Whence,

$$\det(I_d + \Sigma)^{1/d} \leq 1 + \mathrm{Lip}^2(T)\alpha = \det(I_d + \mathrm{Lip}^2(T)\Lambda)^{1/d},$$

as desired. □

Observe that the proof goes through for any contraction $T$ if $G$ is an isotropic Gaussian. This particular case can be generalized to allow for $G$ here to be replaced by an isotropic log-concave distribution $X$ with conditions on $T$ that are directly related to a measure of the distance of $X$ from being Gaussian.

**Theorem 2.3.4.** *Let $X$ be an isotropic log-concave random vector in $\mathbb{R}^d$, $Z$ standard Gaussian. Let $\Delta(X) = \frac{h(Z_X) - h(X)}{d}$, where $Z_X$ is a centered Gaussian in $\mathbb{R}^d$ having the same covariance matrix as $X$. Then, for any contraction $T : \mathbb{R}^d \to \mathbb{R}^d$, we have*

$$N(X+Z) \geq N(T(X)+Z) + (1 - (e^{\Delta(X)}\mathrm{Lip}(T))^2)N(X).$$

*Proof.* Suppose the covariance matrix of $X$ is $\Sigma_X = \alpha I_d$, for some $\alpha > 0$. By the definition of $\Delta(X)$, we have the following equality on the entropy of $X$:

$$h(X) = \frac{d}{2}\log\alpha + \frac{d}{2}\log(2\pi e) - \Delta(X)d.$$

In terms of the entropy power $N(X) = e^{2h(X)/d}$, this reads

$$N(X) = e^{-2\Delta(X)} 2\pi e (\det \Sigma_X)^{1/d} = e^{-2\Delta(X)} 2\pi e \alpha.$$

By entropy power inequality, we have

$$N(X+Z) \geq N(X) + N(Z) \geq 2\pi e (1 + e^{-2\Delta(X)}\alpha).$$

On the other hand, for a Lipschitz map $T$, again by the fact that Gaussian random variables maximize entropy under second-moment constraints, and the AM-GM inequality, one has

$$N(T(X)+Z) \leq 2\pi e (\det(\Sigma_{T(X)} + I_d))^{1/d} \leq 2\pi e (1 + \alpha \mathrm{Lip}^2(T)).$$

The desired inequality now follows. □

Bobkov and Madiman [18, Theorem 2] showed, building upon the work of Keith Ball [8], that the hyperplane conjecture in convex geometry is equivalent to the existence of a constant $C$ such that $\Delta(X) \leq C$ for any log-concave random vector $X$ in any dimension. A crucial observation in [18] is that the isotropic constant of an $\mathbb{R}^d$-valued log-concave random vector $X$, $L_X = \sqrt{h_\infty^{1/d}} \det \sigma_X^{1/2d}$, is related to $\Delta(X)$ via

$$\log\left(\sqrt{\frac{2\pi}{e}} L_X\right) \leq \Delta(X).$$

Therefore, one can also easily write a condition for $T$ in terms of the isotropic constant under which the desired entropic inequality holds.

**Corollary 2.3.5.** *Suppose $X$ is an isotropic log-concave random vector in $\mathbb{R}^d$ with isotropic constant $L_X$. Then for any Lipschitz map $T : \mathbb{R}^d \to \mathbb{R}^d$ with $\mathrm{Lip}(T) \leq \sqrt{\frac{e}{2\pi L_X^2}}$, we have*

$$h(T(X)+Z) \leq h(X+Z).$$

## 2.4 Rényi entropy of order two

A more direct analysis of metric properties of Rényi entropy is sometimes possible if one uses perturbations of Rényi entropies where the metric makes an explicit appearance. The family of perturbation that we use in this section are called diversities. They were introduced in the context of quantification of biodiversity by Leinster and Cobbold [67]. Mathematical aspects of this notion were developed further by Leinster and Meckes [64], Leinster and Roff [66], and Aishwarya, Madiman, Meckes, and the present author [2].

**Definition 8.** Let $(X, d)$ be a metric space and $\mu \in \mathcal{P}(X)$. The diversity of order $\alpha$ of $\mu$ is defined by

$$D_\alpha^K(\mu) = \begin{cases} \left(\int (1/K\mu(x))^{1-\alpha} \ \mathrm{d}\mu(x)\right)^{1/(1-\alpha)} & \text{if } \alpha \in [0,1) \cup (1,\infty), \\ e^{-\int \log K\mu(x) \ \mathrm{d}\mu(x)} & \text{if } \alpha = 1, \\ \frac{1}{\text{ess sup}_\mu K\mu(x)} & \text{if } \alpha = \infty, \end{cases}$$

where $K\mu(x) = \int_X e^{-d(x,y)} \ \mathrm{d}\mu(y)$.

*Remarks.*

1. The logarithms $\log D_\alpha^K$ correspond to Rényi entropy-like quantities.
2. Instead of $e^{-d(x,y)}$ one could use other "kernels" $K(x,y)$. This explains the superscript $K$ of $D_\alpha^K$.
3. If $(X, d)$ is a metric space then so is $(X, td)$ for $t > 0$. We will denote the corresponding diversity measures by $D_\alpha^{K^t}$ or simply $D_\alpha^t$ when there is no risk of confusion.

For a metric structure on a finite set $X$, all integrals in the definition are just finite sums. From here one easily takes limits term-by-term to see that $\lim_{t\to\infty} D_\alpha^t(\mu) = e^{H_\alpha(\mu)}$, where $H_\alpha(\mu) = \frac{1}{1-\alpha} \log \sum_{x\in X} \mu(x)^\alpha$. Thus, we see that all Rényi entropies can be recovered in this case. Here we are interested in the space $X = \mathbb{R}^d$ equipped with the standard Euclidean metric. In this setup too, one can recover Rényi entropies (see [2, Proposition 2.9]). However, all we need is the special case below.

**Lemma 2.4.1.** *Let $\mu$ be a probability measure on $\mathbb{R}^d$ with density with respect to the Lebesgue measure. Then,*

$$\lim_{t\to\infty} C_d \frac{D_2^t(\mu)}{t^d} = e^{h_2(\mu)},$$

*for a constant $C_d$ only depending on the dimension $d$.*

Therefore we get inequalities for the Rényi entropy $h_2$ if we prove the corresponding inequalities for $D_2^t$.

**Theorem 2.4.2.** *Let $X$ be an $\mathbb{R}^d$-valued random vector and $W$ an $\mathbb{R}^d$-valued log-concave random vector with radially-symmetric density. Then, for any contraction $T : \mathbb{R}^d \to \mathbb{R}^d$, $t > 0$, we have*

$$D_2^t(T(X) + W) \leq D_2^t(X + W).$$

*Consequently,*

$$h_2(T(X) + W) \leq h_2(X + W).$$

*Proof.* Note that $D_2^t(Y)$, for any random vector $Y$, can be probabilistically written as

$$D_2^t(Y) = \left(\mathbb{E}_{Y,Y'} e^{-t\|Y-Y'\|}\right)^{-1},$$

where $Y'$ is an independent copy of $Y$. To create an independent copy of $X+W$ we take $X' + W'$, where $X', W'$ are independent copies of $X, W$ respectively and $X, W, X', W'$ are all independent. Now,

$$D_2^t(X + W) = \left(\mathbb{E}_{X,W,X',W'} e^{-t\|(X+W)-(X'+W')\|}\right)^{-1} = \left(\mathbb{E}_{X,X'} \mathbb{E}_{W,W'} e^{-t\|(X-X')-(W'-W)\|}\right)^{-1}.$$

Similarly,

$$D_2^t(T(X) + W) = \left(\mathbb{E}_{X,X'} \mathbb{E}_{W,W'} e^{-t\|(T(X)-T(X'))-(W'-W)\|}\right)^{-1}.$$

Let $g$ denote the density of $W' - W$. Being the density of a convolution of two radially-symmetric log-concave densities (recall here that $-W$ has the same distribution as $W$), it is radially-symmetric and log-concave. Moreover, the function $\phi(z) = e^{-z}, z \in \mathbb{R}^d$, is also radially-symmetric, log-concave, and consequently so is $\phi \star g$. Observe that, for any fixed values $X = x, X' = x'$,

$$\mathbb{E}_{W,W'} e^{-t\|(T(x)-T(x'))-(W'-W)\|} = \phi \star g(T(x) - T(x'))$$
$$\mathbb{E}_{W,W'} e^{-t\|(x-x')-(W'-W)\|} = \phi \star g(x - x').$$

Then, since the value of a radially-symmetric log-concave function is larger for points closer to the origin, from $\|T(x) - T(x')\| \leq \|x - x'\|$ we get

$$\mathbb{E}_{W,W'} e^{-t\|(T(x)-T(x'))-(W'-W)\|} \leq \mathbb{E}_{W,W'} e^{-t\|(x-x')-(W'-W)\|}.$$

Now the desired result follows by taking expectations. □

*Remark.* Here we remind the reader that when $W \sim \text{Uniform}(\mathcal{B})$, the theorem above is trivial since it follows from the intersection conjecture for two balls.

If $X$ is assumed to have a log-concave density then the main result of this section can be extended to a comparison for all Rényi entropies at the cost of a constant which unfortunately blows up at the order 0 thus not giving a direct inequality for volume. The main ingredient is [75, Lemma 2.4] (where the result is attributed to [77]; for a proof one can also see [13, Corollary 4]), which says that for an $\mathbb{R}^d$-valued log-concave random vector all Rényi entropies can be compared via

$$h_\beta(X) - h_\alpha(X) \leq d\left(\frac{\log \beta}{\beta - 1} - \frac{\log \alpha}{\alpha - 1}\right),$$

for $\alpha \geq \beta > 0$. Combined with the fact that Rényi entropies of a fixed random vector are non-increasing as a function of order, we obtain the following corollary.

**Corollary 2.4.3.** *Let $X$ be an $\mathbb{R}^d$-valued log-concave random vector, $W$ an $\mathbb{R}^d$-valued radially-symmetric log-concave random vector, and $T : \mathbb{R}^d \to \mathbb{R}^d$ a contraction. Then,*

$$h_\alpha(T(X) + W) \leq h_\alpha(X + W) + \operatorname{sgn}(2 - \alpha) \left( \frac{\log \alpha}{\alpha - 1} - \log 2 \right) d,$$

*for $\alpha > 0$. In particular,*

$$h(T(X) + W) \leq h(X + W) + 0.307d,$$

*or in other words,*

$$N(T(X) + W) \leq 1.85N(X + W).$$

□

# Chapter 3

# ON THE CENTRAL LIMIT THEOREM OF INFORMATION CONTENT FOR LOG-CONCAVE DENSITIES

## 3.1 Background

Consider the standard Gaussian random vector $Z = (Z_1, \dots, Z_n)$ in $\mathbb{R}^n$, as mentioned in the first chapter, since $\|Z\|^2$ is a sum of i.i.d. random variables, we obtain the so-called Chernoff's bound:

$$\mathbb{P}\left\{\frac{\|Z\|^2}{n} - 1 > t\right\} \leq \exp\left\{-\frac{n}{2}[t - \log(1+t)]\right\}. \tag{3.1}$$

Fradelizi, Madiman, and Wang [43] have shown that the concentration inequality (3.1) has an extension to all log-concave random vectors in $\mathbb{R}^n$:

$$\mathbb{P}\left\{-\log f(X) + \mathbb{E}\log f(X) > nt\right\} \leq \exp\left\{-nr(t)\right\}, \tag{3.2}$$

$$\mathbb{P}\left\{-\log f(X) + \mathbb{E}\log f(X) < -nt\right\} \leq \exp\left\{-nr(-t)\right\}, \tag{3.3}$$

where $r(t) = t - \log(1+t)$ for $t > -1$, and $r(t) = +\infty$, otherwise.

The concentration inequalities (3.2) and (3.3) follows from the sharp variance bound

$$V(X) := \mathrm{Var}(-\log f(X)) \leq n, \tag{3.4}$$

which was first established by V.H. Nguyen [85] and L. Wang [103]. As a consequence of inequality (3.3), one deduces the following entropy comparison inequality, which was first explicitly proved in [15], also implicitly contained in the earlier work (see, e.g., [45]),

$$\mathbb{E}[-\log f(X)] \leq -\log \|f\|_\infty + n. \tag{3.5}$$

It is an open problem to characterize the equality case for inequality (3.4). In the sequel, we will characterize the equality case in (3.5), which, in turns, implies the equality case of (3.4).

Note that, since $\|Z\|^2$ is a sum of i.i.d. random variables, we may use the Berry-Esseen theorem to obtain

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{\|Z\|^2 - n}{\sqrt{2n}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \,\mathrm{d}r \right| \leq \frac{c}{\sqrt{n}}, \tag{3.6}$$

where $c$ is an absolute constant. Meanwhile, we also observe that inequality (3.6) can be rewritten in terms of the density of $Z$ as follows:

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{-\log \phi(Z) + \mathbb{E}(\log \phi(Z))}{\sqrt{\mathrm{Var}(-\log \phi(Z))}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \,\mathrm{d}r \right| \leq \frac{c}{\sqrt{n}}, \tag{3.7}$$

where $\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\|x\|^2}{2}}$.

We conjectured that (3.7) can also be generalized to any $\mathbb{R}^n$-valued random variable $X$ with log-concave density $f(x) = e^{-U(x)}$ such that $\mathrm{Var}(-\log f(X)) = \beta n$ for some absolute constant $\beta > 0$, i.e.

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{-\log f(X) + \mathbb{E}(\log f(X))}{\sqrt{\mathrm{Var}(-\log f(X))}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \,\mathrm{d}r \right| \leq \frac{c}{\sqrt{n}}, \tag{3.8}$$

where $c$ is an absolute constant that only depends on $\beta$.

Interestingly, it turns out that the above conjecture is indeed true for log-concave random vector $X$ satisfying the equality case in (3.5). Motivated by this result, we show that the conjecture holds for a subclass of log-concave random vectors, which directly generalizes the Gaussian case (3.7). Meanwhile, a qualitative central limit theorem conjecture is also formulated in Conjecture 9,and the qualitative central limit theorem conjecture can be fully settled if one is able to prove a uniform bound for the third moments of a family of random variables (see Theorem 3.3.3 for more details), which is an observation that goes back to L. Wang [103].

After introducing the necessary background, we start with the characterization of the equality case in (3.5).

### 3.2 Characterization of convex potentials

**Definition 9.** Let $X$ be a random vector in $\mathbb{R}^n$ with probability density function $f$. The information content of $X$ is the random variable $\tilde{h}(X) := -\log f(X)$. The entropy of $X$ is defined as $h(X) = \mathbb{E}(-\log f(X))$. The varentropy $V(X)$ of $X$ is defined as $V(X) = \mathrm{Var}(-\log f(X))$.

*Remark.* Given a random vector $X$ with log-concave density $f(x)$. Let $T$ be an invertible affine transformation, then $TX$ is a random variable having density

$$g(x) = |\det T|^{-1} f(T^{-1}x).$$

Therefore

$$-\log g(TX) + \mathbb{E}\log g(TX) = -\log f(X) + \mathbb{E}\log f(X).$$

Hence the varentropy $V(X)$ is affine invariant.

Recall that a random variable $X$ taking values in $\mathbb{R}^n$ is said to be log-concave if it has a log-concave density $f(x) = e^{-U(x)}$. The convex function $U(x)$ is called the potential function of $f(x)$. A convex potential function $U(x)$ is said to be positively homogeneous of degree 1, if $U(tx) = tU(x)$ for all $t > 0$ and $x \in \mathbb{R}^n$. Given a random variable $X$ in $\mathbb{R}^n$ with density $f(x)$, let $X^*$ denote the random variable with density $f^*(x)$, the symmetrically decreasing rearrangement of $f(x)$.

There are various ways of strengthening log-concavity, one of which is due to Bobkov and Madiman [14], where the log-concave distribution of order $p \geq 1$ is introduced.

**Definition 10.** A function $f(x)$ from $[0,\infty)$ to $[0,\infty)$ is said to be log-concave of order $p \geq 1$, if

$$f(x) = x^{p-1} g(x),$$

where $g(x)$ is a log-concave function from $[0,\infty)$ to $[0,\infty)$. A probability measure is said to be log-concave of order $p$, if it has a density function that is log-concave of order $p$.

Log-concavity is a natural shape constraint for probability density functions that represents an infinite dimensional generalization of the class of Gaussian distributions. With log-concavity, it turns out that one can obtain sharp upper bounds for both the varentropy $V(X)$ and entropy $h(X)$.

**Theorem 3.2.1.** *[43] Let $X$ be a log-concave random vector in $\mathbb{R}^n$ with density $f = e^{-U(x)}$, then*

$$V(X) \leq n,$$

$$h(X) \leq -\log \|f\|_\infty + n.$$

*Remark.* These upper bounds are sharp and attained when $U(x)$ is positively homogeneous of degree 1, which, in turn, implies that $\|f\|_\infty = 1$. Indeed, positive homogeneity of $U(x)$ implies that $U(0) = 0$, meanwhile, since $f$ is a log-concave density function, there exist constant $A, B > 0$ such that $f(x) \leq Ae^{-B\|x\|_2}$ for all $x \in \mathbb{R}^n$ (see, e.g., [6, Lemma 10.6.1]). Therefore, it is necessary that $U(x) \geq U(0) = 0$ for all $x \in \mathbb{R}^n$. For more details, please see Theorem 3.2.2 below.

Conversely, for any $\mathbb{R}^n$ valued log-concave random vector which achieves the entropy upper bound, it turns out that, up to an affine transformation, its potential function is positively homogeneous of degree 1. In fact, we have the following theorem.

**Theorem 3.2.2.** *Let $X$ be a log-concave random vector in $\mathbb{R}^n$ with law $\mu$ and density $f(x) = e^{-U(x)}$, so that $\min_x U(x) = U(0) = 0$; and let $\nu = U\#\mu$ be the push forward measure of $\mu$ by $U(x)$. The the following statements are equivalent:*

1. *$U(x)$ is positively homogeneous of degree 1.*

*2.* $h(X) = n$.

*3.* $h(X^*) = n$.

*4.* $V(X) = n$, *and* $\nu$ *is log-concave of order* $n$.

*5.* $V(X_\alpha) = n$ *on a set of* $\alpha$ *with at least one accumulation point.*

*6. For* $\beta < 1$,
$$\mathbb{E}e^{\beta[U(X)-\mathbb{E}U(X)]} = e^{-n[\beta+\log(1-\beta)]}.$$

To prove our theorem, we need some preliminaries. Convex functions play an prominent role in our work, we therefore first collect a few facts about them. Most of those facts or properties can be found in [92]. Recall that the domain of a convex function $U(x)$ in $\mathbb{R}^n$ is defined as the set

$$\text{Dom}(U) := \{x \in \mathbb{R}^n : U(x) < \infty\}.$$

Since $U(x)$ is convex, we have that Dom$(U)$ is a convex set, the boundary of which, therefore, has Lebesgue measure 0. Let Int(Dom$(U)$) be the largest open set contained in Dom$(U)$. It is well known (see, e.g. , [92]) that convex functions are continuous and locally Lipschitz on Int(Dom(U)). As a consequence, by Rademacher's theorem, $\nabla U$ is well-defined almost everywhere and locally bounded (see, e.g. , [39, subsection 3.1.2]). For those points in Int(Dom(U)) where $U$ is not differentiable, we could define the sub-differential of $U$ at the non-differentiable point $x \in$ Int(Dom(U)) as follows:

$$\partial U(x) = \{v \in \mathbb{R}^n : U(z) - U(x) \geq \langle v, z - x\rangle,\ \forall z \in \mathbb{R}^n\}.$$

In the sequel, if $U(x)$ is not differentiable at $x \in$ Int(Dom(U)), we will pick an arbitrary element from $\partial U(x)$, and denote it by $\nabla U(x)$.

Since one may modify the values of $U(x)$ on the boundary of Dom$(U)$ without harming the convexity property of $U(x)$, we have the following modification lemma.

**Lemma 3.2.3.** *Let* $f(x) = e^{-U(x)}$ *be a log-concave density function such that* $U(0) = 0$ *and* $U(x)$ *is positively homogeneous of degree 1 on Int(Dom(*$U$*)). Then* $U(x)$ *can be* *modified to a convex function* $\tilde{U}(x)$ *such that* $\tilde{U}(x)$ *is positively homogeneous of degree 1 on* $\mathbb{R}^n$ *and* $\tilde{f}(x) = e^{-\tilde{U}(x)}$ *is again a probability density.*

*Proof.* If $0 \in \text{Int}(\text{Dom}(U))$, then 1-homogeneity implies that $\text{Int}(\text{Dom}(U)) = \mathbb{R}^n$ and the modification is clearly $U(x)$ itself. Therefore, we may assume that $0 \in \partial\text{Dom}(U)$. Let $L$ the smallest open convex cone centered at the origin such that $\text{Int}(\text{Dom}(U)) \subset L$. Since $\text{Int}(\text{Dom}(U))$ is also a convex set, 1-homogeneity of $U(x)$ on $\text{Int}(\text{Dom}(U))$ implies that $\text{Int}(\text{Dom}(U)) = L$. Therefore, we may define $\tilde{U}(x) = U(x)$ on $\text{Int}(\text{Dom}(U))$, and set $\tilde{U}(x) = +\infty$, otherwise. Now one can easily verify that $\tilde{U}(x)$ satisfies all the requirements. $\square$

In the rest of the thesis, a log-concave density function $f(x) = e^{-U(x)}$ is said to have a positive 1-homogeneous potential if $U(x)$ is positively homogeneous of degree 1 on $\text{Int}(\text{Dom}(U))$.

Another ingredient in our work is the spherically decreasing rearrangement. Rearrangement theory is about reorganizing functions or sets in some specific way. It is an analytic tool which plays a crucial role in geometry and integration theory. For instance, from the geometric point of view, rearrangement theory leads to the Brunn-Minkowski inequality in both continuous (see, e.g., [6, subsection 1.2.1]) and discrete setting [49]. It is also used to prove the Pólya-Szegó inequality (see, e.g., [26]), which, in turn, shows that among all open subsets of $\mathbb{R}^n$ with fixed positive Lebesgue measure, the Laplace operator $-\Delta$ defined on the open ball (with Dirichlet boundary condition) bears the smallest first eigenvalue (see, e.g., [51, section 4]). From the functional point of view, the rearrangement method is used to determine the sharp constant for Young's inequality, which unifies the Brunn-Minkowski inequality and entropy power inequality [21,69]; and it also unveils the extremizer for Sobolev's inequality (see, e.g., [70, Chapter 8]), which can be seen as a functional lift of the isoperimetric inequality. The reader is referred to the beautiful survey paper written by Gardner [48] for more information.

In our paper, we will focus on the spherically decreasing rearrangement, recall the Definition 6 given in the second chapter. For any Borel measurable $f : \mathbb{R}^d \to [0, \infty)$

which vanishes at infinity, we define its symmetrically-decreasing rearrangement as follows:

$$f^*(x) = \int_0^\infty \mathbf{1}_{\{y:|f(y)|>t\}^*}(x)\ \mathrm{d}t.$$

*Remarks.*

1. One can describe the super level sets of $f^*$ explicitly by definition

$$\{x : f^*(x) > t\} = \{x : |f(x)| > t\}^*, \text{ for every } t > 0. \tag{3.9}$$

   In particular, this implies the equi-measurability of $|f|$ and $f$:

$$\mathcal{L}^n(\{x : f^*(x) > t\}) = \mathcal{L}^n(\{x : |f(x)| > t\}), \text{ for every } t > 0, \tag{3.10}$$

   where $\mathcal{L}^n$ represents the Lebesgue measure in $\mathbb{R}^n$.

2. As terminology indicates, $f^*$ is indeed radially-symmetric and decreases radially.

For more information regarding rearrangements, we refer to Lieb and Loss' text [70, Chapter 3] and Burchard's notes [26].

The following lemma, regarding the spherically decreasing rearrangement, says that, for any log-concave density function $f$, the log-concavity is preserved by the spherically decreasing rearrangement.

**Lemma 3.2.4.** *[81, Corollary 5.4] Suppose $X$ is a log-concave random vector taking value in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$. Let $f^*$ be the symmetric decreasing rearrangement of $f$. Then $f^*$ is also log-concave.*

*Remarks.* As a consequence of equi-measurability (3.10) and layer cake representation (see, e.g., [70, Theorem 1.13]), one has

- For $\alpha \geq 1$, if $f \in L^\alpha(\mathbb{R}^n)$, so is $f^*$, and $\|f\|_\alpha = \|f^*\|_\alpha$. In particular, for log-concave density $f$, $f^*$ is again a log-concave probability density function by the previous lemma; and therefore, the Rényi entropy (for definition, see, e.g. [75]) $h_\alpha(f) = h_\alpha(f^*) < \infty$ for all $\alpha > 1$, and we may now send $\alpha \searrow 1$ by continuity to deduce that $h(f) = h(f^*)$.

- Denote by $X^*$ the $\mathbb{R}^n$ valued random vector with density $f^*$, we have $V(X) = V(X^*)$.

Apart from the properties list above, the following proposition says that the 1-homogeneity of the potential function remains invariant under the action of spherically decreasing rearrangement.

A log-concave function $f(x)$ is said to be positively log-homogeneous of degree 1 if $-\log f(x)$ is positively homogeneous of degree 1.

**Proposition 3.2.5.** *Assume that $f = e^{-U(x)}$ is a log-concave integrable function on $\mathbb{R}^n$ such that $\min_{x\in\mathbb{R}^n} U(x) = U(0) = 0$. Then $f(x)$ is positively log-homogeneous of degree 1 if and only if $f^*$ is.*

*Proof.* If $f$ is positively log-homogeneous of degree 1, then there exists a positively homogeneous convex function $U(x)$ of degree 1, such that $f(x) = e^{-U(x)}$. By definition of spherical decreasing rearrangement, we have

$$f^*(x) = \int_0^\infty \mathbb{1}_{\{f>t\}^*}(x)\ \mathrm{d}t = \int_0^1 \mathbb{1}_{\{U<-\log t\}^*}(x)\ \mathrm{d}t$$

Note that $-\log t > 0$ for $t \in (0,1)$, use positive homogeneity of $U(x)$, we have

$$\{y : U(y) < -\log t\}^* = \{y : \|y\|_2 < -\frac{1}{c_n}\log t\} = \{y : e^{-c_n\|y\|_2} > t\},$$

where $c_n = (\frac{|B_2|}{|K|})^{\frac{1}{n}}$, and $K = \{x : U(x) < 1.\}$

Therefore,

$$f^*(x) = e^{-c_n\|x\|_2}.$$

Conversely, assume that $f^*$ is log-homogeneous of degree 1. For any $r > 0$, by equation (3.10), we have, for all $t > 0$,

$$\frac{|\{x : f^*(x) > e^{-rt}\}|}{t^n} = \frac{|\{x : f(x) > e^{-rt}\}|}{t^n} = \frac{|\{x : U(x) < rt\}|}{t^n}.$$

By homogeneity of Lebesgue measure, we have

$$\frac{|\{x : U(x) < rt\}|}{t^n} = |\{x : \frac{U(tx)}{t} < r\}|.$$

Since $f^*$ is log-homogeneous of degree 1 , we have

$$\frac{|\{x : f^*(x) > e^{-rt}\}|}{t^n} = |\{x : f^*(x) > e^{-r}\}| = |\{x : f(x) > e^{-r}\}| = |\{x : U(x) < r\}|.$$

Therefore, for any $r > 0$ and $t > 0$, we have

$$|\{x : \frac{U(tx)}{t} < r\}| = |\{x : U(x) < r\}| < \infty.$$

For fixed $t > 0$, by layer cake representation, we have

$$\int_{\mathbb{R}^n} e^{-\frac{U(tx)}{t}} \, \mathrm{d}x = \int_{\mathbb{R}^n} e^{-U(x)} \, \mathrm{d}x. \tag{3.11}$$

Using $U(0) = 0$ and the convexity of $U(x)$, for $\alpha > 1$ and arbitrary $x$, we have

$$\alpha U(x/\alpha) \leq U(x).$$

Combining with equation (3.11), we have

$$\alpha U(x/\alpha) = U(x) \quad a.e.$$

Now, use the continuity of $U(x)$ on $\mathrm{Int}(\mathrm{Dom}(U))$ to deduce that $\alpha U(x/\alpha) = U(x)$ for all $x \in \mathrm{Int}(\mathrm{Dom}(U))$.

If $\alpha \in (0, 1)$, we have

$$\alpha U(x/\alpha) \geq U(x),$$

and we can repeat the argument to conclude that

$$\alpha U(x/\alpha) = U(x) \quad \forall \alpha > 0, \quad \forall x \in \mathrm{Int}(\mathrm{Dom}(U)).$$

□

An integrable log-concave function $f(x)$ decays fast enough to guarantee some analytic properties like below.

**Lemma 3.2.6.** *Let $f(x) = e^{-U(x)}$ be a log-concave density function on $\mathbb{R}^n$, and define $T(z) = \int f^z(x)\,\mathrm{d}x$ , then $T(z)$ is holomorphic on $\Omega = \{z \in \mathbb{C} : Re(z) > 0\}$. In particular, $F(\alpha) := \log T(\alpha)$ is real analytic on $\alpha > 0$.*

*Proof.* Let $m \in \mathbb{N}$ and $B_m(0)$ be the ball in $\mathbb{R}^n$ with radius $m$. Define

$$T_m(z) = \int_{B_m(0)} e^{-zU(x)}\,\mathrm{d}x$$

To prove that $T(z)$ is holomorphic on $\Omega$, it suffices to prove that $T_m(z)$ is holomorphic on $\Omega$ and for any closed disc $D \subseteq \Omega$, $T_m(z)$ converges uniformly to $T(z)$.
To this end, fix $m$, for $0 \le k \le 2^l l - 1$, let $A_{k,l,m} := \{x \in B_m(0) : k/2^l \le U(x) < (k+1)/2^l\}$, and $A_{l,l,m} := \{x \in B_m(0) : U(x) \ge l\}$. Define

$$T_m^{(l)}(z) = \sum_{k=0}^{2^l l} e^{-\frac{k}{2^l}z} |A_{k,l,m}|,$$

where $|A|$ represents the Lebesgue measure of the measurable set $A$. It is clear that $T_m^{(l)}(z)$ is holomorphic.
Fix any closed disc $D \subseteq \Omega$,

$$|T_m^{(l)}(z) - T_m(z)| \le \sum_{k=0}^{2^l l} \int_{A_{k,l,m}} |e^{-zU(x)} - e^{-\frac{k}{2^l}z}|\,\mathrm{d}x. \tag{3.12}$$

Let $\alpha_0 = \min_{z\in D} \mathrm{Re}(z) > 0$, and $\beta_0 = \max_{z\in D} |z|$. Note that

$$|e^{-zU(x)} - e^{-\frac{k}{2^l}z}| = \left|\int_{k/2^l}^{U(x)} ze^{-zs}ds\right| \leq \int_{k/2^l}^{U(x)} |ze^{-zs}|ds \leq \frac{\beta_0 e^{\alpha_0}}{2^l} e^{-\alpha_0 U(x)}. \tag{3.13}$$

Combining (1) and (2), we obtain

$$|T_m^{(l)}(z) - T_m(z)| \leq \frac{\beta_0 e^{\alpha_0}}{2^l} \int_{B_m(0)} e^{-\alpha_0 U(x)}dx \leq \frac{\beta_0 e^{\alpha_0}}{2^l} \int e^{-\alpha_0 U(x)}dx$$

which goes to 0 uniformly on $D$ as $l \to \infty$. Therefore $T_m(Z)$ is holomorphic on $\Omega$. Now for the same disc $D$, we have

$$|T_m(z) - T(z)| \leq \int_{\mathbb{R}^n\setminus B_m(0)} |e^{-zU(x)}|dx \leq \int_{\mathbb{R}^n\setminus B_m(0)} e^{-\alpha_0 U(x)}dx,$$

which goes to 0 uniformly on $D$ as $m \to \infty$. The proof is complete.

□

*Remark.* We can also show that $F(\alpha)$ is real analytic on $\alpha > 0$ directly. Indeed, on $\alpha > 0$, $T(\alpha) > 0$ and $\log \alpha$ is real analytic. It therefore suffices to show that $T(\alpha)$ is real analytic when $\alpha > 0$. Note that for any $\alpha > 0$, since $f$ is a log-concave density function, one has that $f \in L^\alpha(\mathbb{R}^n)$. Given an arbitrary compact set $K \subseteq \mathbb{R}^+$, let $\alpha_0 = \min_{\alpha\in K} \alpha$. For $n \in \mathbb{N}$,

$$\frac{|T^{(n)}(\alpha)|}{n!} \leq \int e^{-\alpha U(x)} \frac{U(x)^n}{n!}dx \leq \int e^{-\alpha_0 U(x)} \frac{U(x)^n}{n!}dx \leq \left(\frac{2}{\alpha_0}\right)^n \int e^{-\frac{\alpha_0 U(x)}{2}} \,\mathrm{d}x \leq (C_{\alpha_0})^n,$$

from which the analyticity of $T(\alpha)$ follows.

For a given log-concave random vector $X$ taking values in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$, $F(\alpha)$ is actually the log-moment generating function of the random variable $U(X)$. To see its connection with varentropy $V(X)$, we define a family of $\mathbb{R}^n$

valued random vectors $X_\alpha$ with densities

$$f_\alpha := \frac{f^\alpha}{\int_{\mathbb{R}^n} f^\alpha \, \mathrm{d}x}.$$

A direct calculation reveals that

$$\alpha^2 F''(\alpha) = V(X_\alpha). \tag{3.14}$$

It might be not too surprising to connect the derivatives of log-moment generating function with the centered moment of the tilted random variables, especially for those readers who are familiar with large deviation principles. We also want to emphasize that although $V(X_\alpha)$ is real analytic by (3.14) and bounded above by $n$ (see Theorem 3.2.1), $V(X) = n$ does not imply that $V(X_\alpha) = n$ for all $\alpha > 0$, since real analytic functions do not enjoy the property of maximum principle. However, log-concavity allows us to deduce the third moment of $U(X) - \mathbb{E}(U(X))$ when the second moment, i.e, the varentropy, is saturated. To be more precise, we have the following proposition,

**Proposition 3.2.7.** *Suppose that $X$ is a log-concave random vector taking values in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$. If $V(X) = n$, then we have*

$$K_3(X) := \mathbb{E}[U(X) - \mathbb{E}(U(X))]^3 = 2n.$$

*Proof.* By [43, Theorem 2.9], one has that $G(\alpha) := \alpha^n \int f^\alpha(x) \, \mathrm{d}x$ is a log-concave function of $\alpha$ for $\alpha > 0$. We will exploit this result a little bit deeper. Recall that $F(\alpha)$ is defined as follows,

$$F(\alpha) = \log \int f^\alpha.$$

and $F(\alpha)$ is real analytic. Since $G(\alpha)$ is log-concave, $\log G(\alpha) = n \log \alpha + F(\alpha)$ is an infinitely differentiable concave function. Define

$$H(\alpha) := (\log G(\alpha))' = \frac{n}{\alpha} + F'(\alpha),$$

which is also infinitely differentiable. We have, for $\alpha > 0$,

$$H'(\alpha) = -\frac{n}{\alpha^2} + F''(\alpha) \leq 0,$$

because of the concavity of $\log G(\alpha)$. A direct calculation shows that $F''(\alpha) = \frac{V(X_\alpha)}{\alpha^2}$ and $V(X) = F''(1)$. Therefore, if $V(X) = n$, we have

$$H'(1) = 0.$$

Since $H'(\alpha) \leq 0$ for $\alpha > 0$. we have the following

$$\lim_{\alpha\to 1^-} \frac{H'(\alpha) - H'(1)}{\alpha - 1} = \lim_{\alpha\to 1^-} \frac{H'(\alpha)}{\alpha - 1} \geq 0.$$

$$\lim_{\alpha\to 1^+} \frac{H'(\alpha) - H'(1)}{\alpha - 1} = \lim_{\alpha\to 1^+} \frac{H'(\alpha)}{\alpha - 1} \leq 0.$$

Since $H(\alpha)$ is infinitely differentiable, it forces that

$$H''(1) = 0,$$

while

$$H''(\alpha) = \frac{2n}{\alpha^3} + F'''(\alpha) = \frac{2n}{\alpha^3} - \frac{K_3(X_\alpha)}{\alpha^3}.$$

We therefohave

$$H''(1) = 2n - K_3(X) = 0,$$

as desired.

$\square$

Since $f_\alpha$ is again log-concave for any $\alpha > 0$, by Theorem 3.2.1, one has $V(X_\alpha) \leq n$, for all $\alpha > 0$. The following lemma says that if $V(X_\alpha) = n$ for all $\alpha > 0$, then it is necessary that the potential function $U(x)$, up to an affine transformation, is positively homogeneous of degree 1.

**Lemma 3.2.8.** *Suppose $X$ is a log-concave random vector in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$ such that $\min_x U(x) = U(0) = 0$, and $V(X_\alpha) = n$ for all $\alpha > 0$. Then $U(x)$ is a positively homogeneous function of degree 1.*

*Proof.* Indeed, under this condition, we have $G(\alpha)$ is a log-affine function. Since $G(1) = 1$, therefore,

$$\log G(\alpha) = c(\alpha - 1),$$

for some constant $c$.

On the other hand,

$$\log G(\alpha) = n \log \alpha + F(\alpha),$$

whence we have

$$c = \frac{n}{\alpha} + F'(\alpha).$$

Setting $\alpha = 1$, we obtain $c = n - H(X)$.

By [43, Corollary 4.5], we have

$$H(X) \leq -\log \|f\|_\infty + n.$$

Therefore $c \geq 0$. If $c > 0$, we have

$$\lim_{\alpha \to \infty} \frac{F(\alpha)}{\alpha} = c.$$

Thus we have

$$\int f^\alpha(x) dx = e^{\alpha c + o(\alpha)}$$

However, for $\alpha$ large enough,

$$|F'(\alpha)| = \frac{\int e^{-\alpha U(x)} U(x)\ \mathrm{d}x}{\int f^{\alpha}(x)\ \mathrm{d}x} \leq \frac{\int f^{\frac{\alpha}{2}}(x)\ \mathrm{d}x}{\int f^{\alpha}(x)\ \mathrm{d}x} = e^{-\frac{\alpha c}{2} + o(\alpha)} \to 0$$

Let $\alpha \to \infty$, we have $c = 0$, which is a contradiction!

Therefore, we must have $c = 0$, and

$$\int e^{-\alpha U(x)}\ \mathrm{d}x = \alpha^{-n} \Longrightarrow \int e^{-\alpha U(x/\alpha)}\ \mathrm{d}x = 1. \tag{3.15}$$

But since $f(x) = e^{-U(x)}$ is a probability density, we have

$$\int \left(e^{-\alpha U(x/\alpha)} - e^{-U(x)}\right)\ \mathrm{d}x = 0.$$

Now repeat the argument in Proposition 3.2.5 to conclude that

$$\alpha U(x/\alpha) = U(x) \quad \forall \alpha > 0, \quad \forall x \in \mathrm{Int}(\mathrm{Dom}(U)).$$

□

The last ingredient for proving our main theorem in this section will be an auxiliary inequality due to Borell [20].

**Theorem 3.2.9.** *Let $f : [0, \infty) \to [0, \infty)$ be a log-concave function. Then, the function*

$$G(p) := \frac{1}{\Gamma(p+1)} \int_0^{\infty} x^p f(x)\ \mathrm{d}x$$

*is a log-concave function of $p$ on $[0, \infty)$.*

Finally, we are ready to prove our main Theorem 3.2.2 in this section.

*Proof of Theorem 3.2.2.*

- $1 \Rightarrow$ the rest. Indeed, if a log-concave random vector $X$ has density $f(x) = e^{-U(x)}$ with the potential function $U(x)$ being positively homogeneous of degree 1, i.e., $U(tx) = tU(x)$ for all $t > 0$ and $x \in \mathbb{R}^n$, then the Laplace transform of the random variable $U(X)$ can be computed as follows: for any $t > -1$,

$$\mathbb{E}e^{-tU(X)} = \int_{\mathbb{R}^n} e^{-(1+t)U(x)}\,\mathrm{d}x = \int_{\mathbb{R}^n} e^{-U((1+t)x)}\,\mathrm{d}x = (1+t)^{-n}.$$

Therefore, we deduce that

$$h(X) = \mathbb{E}(U(X)) = n.$$

Hence, for $\beta < 1$, we have

$$\mathbb{E}e^{\beta[U(X)-\mathbb{E}U(X)]} = e^{-n[\beta+\log(1-\beta)]},$$

which, in turn, implies that

$$V(X) = \mathrm{Var}(U(X)) = n.$$

For $\alpha > 0$, set $t = \alpha - 1$, and recall the definition for $F(\alpha)$,

$$F(\alpha) = \log \int_{\mathbb{R}^n} f^{\alpha}(x)\,\mathrm{d}x = -n\log\alpha.$$

By equation (3.14), we have, for $\forall \alpha > 0$,

$$V(X_\alpha) = n.$$

Moreover, the Laplace transform of $U(X)$ reveals that

$$U(X) \stackrel{d}{=} \mathrm{Gamma}(n, 1),$$

where the Gamma$(n,1)$ has the gamma distribution with density $f(t) = \frac{1}{\Gamma(n)} t^{n-1} e^{-t}$ for $t \geq 0$. Therefore, $\nu$ is log-concave of order $n$. The equivalence of 2 and 3 follows from the remarks after Lemma 3.2.4 and Proposition 3.2.5.

- $2 \Rightarrow 1$. Let us first assume that $\mathrm{Dom}(U) = \mathbb{R}^n$. Using the convexity of $U(x)$ and $U(0) = 0$, we have

$$U(x) \leq \langle x, \nabla U(x) \rangle. \tag{3.16}$$

Therefore,

$$n = h(X) = \int_{\mathbb{R}^n} U(x)e^{-U(x)}\ \mathrm{d}x \le \int_{\mathbb{R}^n} \langle x, \nabla U(x)\rangle e^{-U(x)}\ \mathrm{d}x = -\int_{\mathbb{R}^n} x \cdot \nabla e^{-U(x)}\ \mathrm{d}x. \tag{3.17}$$

Since $f(x)$ is a log-concave density function, there exist constant $A, B > 0$ such that $f(x) \le Ae^{-B\|x\|_2}$ for all $x \in \mathbb{R}^n$. Now a direct application of the divergence theorem reveals that

$$\int_{\mathbb{R}^n} x \cdot \nabla e^{-U(x)}\ \mathrm{d}x = -n. \tag{3.18}$$

Combining (3.16), (3.17), (3.18), and the continuity of $U(x)$, we have, for every $x \in \mathbb{R}^n$,

$$U(x) = \langle x, \nabla U(x)\rangle.$$

For a fixed $x$, $U(tx)$ is a convex function on $\mathbb{R}^+$, which is differentiable almost everywhere. Moreover, when it is differentiable, we have

$$\frac{\partial U(tx)}{\partial t} = \langle \nabla U(tx), x\rangle = \frac{U(tx)}{t}.$$

Solving the differential equation(recall that $U(0) = 0$), and using continuity of $U(x)$ again, we arrive at

$$U(tx) = tU(x), \quad \forall t \ge 0.$$

This implies that $U(x)$ is positively homogeneous of degree 1.

Now assume that $\mathrm{Dom}(U) \ne \mathbb{R}^n$. We claim that $\mathrm{Dom}(U)$ cannot be a bounded convex set of $\mathbb{R}^n$.

Indeed, if $\mathrm{Dom}(U)$ is bounded, then $\overline{\mathrm{Dom}(U)}$ is a compact convex set (i.e., a convex body). For any $\theta \in \mathbb{S}^{n-1}$, denote by $h_K(\theta)$ the support function (for definition, see e.g., [6]) of a convex body $K$. Since $\partial\mathrm{Dom}(U)$ has Lebesgue measure 0, and those points in $\mathrm{int}(\mathrm{Dom}(U))$ where $U$ is not differentiable also have Lebesgue measure 0, we may repeat the previous argument to get that, by (3.16),

$$n = \int_{\mathbb{R}^n} U(x)e^{-U(x)}dx = \int_{\mathrm{int}(\mathrm{Dom}(U))} U(x)e^{-U(x)}dx$$

$$\le \int_{\mathrm{int}(\mathrm{Dom}(U))} \langle x, \nabla U(x)\rangle e^{-U(x)}dx = -\int_{\mathrm{int}(\mathrm{Dom}(U))} x \cdot \nabla e^{-U(x)}dx, \tag{3.19}$$

By the divergence theorem, we have

$$-\int_{\overline{\mathrm{Dom}(U)}} x \cdot \nabla e^{-U(x)} dx = n - \int_{\partial \mathrm{Dom}(U)} x e^{-U(x)} \cdot \ d\vec{S}$$

$$= n - \int_{\partial \mathrm{Dom}(U)} e^{-U(x)} \langle x, n_x \rangle \ d\sigma(x), \tag{3.20}$$

where $d\sigma(x)$ is the usual surface area measure, and $n_x$ is the outer normal vector of the convex body $K_m$ at $x$.

Since Dom(U) is convex and $o \in \mathrm{Dom}(U)$, we have $\langle x, n_x \rangle = h_{\overline{\mathrm{Dom}(U)}}(n_x) \geq 0$, for all $x \in \partial \mathrm{Dom}(U)$. Therefore, inequality (3.19) and equality (3.20) imply that

$$\int_{\partial \mathrm{Dom}(U)} e^{-U(x)} h_{\overline{\mathrm{Dom}(U)}}(n_x) \ d\sigma(x) = 0, \tag{3.21}$$

and as a consequence,

$$U(x) = \langle x, \nabla U(x) \rangle, \ \forall x \in \mathrm{int}(\mathrm{Dom}(U)).$$

Since $o \in \mathrm{Dom}(U)$, one may replace $\mathbb{R}^n$ with $\mathrm{int}(\mathrm{Dom}(U))$ in the previous proof. Then we have, for any $x \in \mathrm{int}(\mathrm{Dom}(U))$, there exists $t_x > 0$ such that

$$U(tx) = tU(x), \ \forall t \in [0, t_x). \tag{3.22}$$

If $\mathrm{Dom}(U)$ is bounded, on one hand ,we have $t_x < \infty$ for arbitrary $x \in \mathrm{int}(\mathrm{Dom}(U))$, which implies that we may assume $U(x)$ is finite on $\partial \mathrm{Dom}(U)$. On the other hand, boundedness of $\mathrm{Dom}(U)$ enables us to find a measurable set $A \subseteq \partial \mathrm{Dom}(U)$ with $\sigma(A) > 0$, such that $h_{\overline{\mathrm{Dom}(U)}}(n_x) > 0$ for all $x \in A$. But this contradicts equation (3.21) and the claim follows.

By approximating the unbounded convex set $\overline{\mathrm{Dom}(U)}$ with compact convex sets, and repeating the argument above, one can again get (3.22). To extend equation (3.22) to all $t > 0$, we will show that $\mathrm{int}(\mathrm{Dom}(U))$ is an open convex cone centered at the origin. Indeed, if not, then there always exists $r > 0$, such that the following are true:

- $\sigma(\mathcal{B}_r(0) \bigcap \partial \mathrm{Dom}(U)) > 0$.
- there exists a set $A \subset \mathcal{B}_r(0) \bigcap \partial \mathrm{Dom}(U)$ such that $\sigma(A) > 0$, and $h_{\overline{\mathrm{Dom}(U)}}(n_x) > 0$, for all $x \in A$.

Now choose a sequence of compact convex sets $K_m$, the boundary of which contain $\mathcal{B}_r(0) \bigcap \partial \mathrm{Dom}(U)$, such that $K_m \nearrow \overline{\mathrm{Dom}(U)}$. Then

$$n = \int_{\mathbb{R}^n} U(x) e^{-U(x)} \ \mathrm{d}x = \int_{\mathrm{Dom}(U)} U(x) e^{-U(x)} \ \mathrm{d}x = \lim_{m \to \infty} \int_{K_m} U(x) e^{-U(x)} \ \mathrm{d}x.$$

Note that,

$$\int_{K_m} U(x)e^{-U(x)}\ \mathrm{d}x \leq \int_{K_m} \langle x, \nabla U(x)\rangle e^{-U(x)}\ \mathrm{d}x = -\int_{K_m} x\cdot\nabla e^{-U(x)}\ \mathrm{d}x.$$

Since $0 \in \partial\mathrm{Dom}(U)$, this implies that $h_{\overline{\mathrm{Dom}(U)}}(n_x) \geq 0$, for all $x \in \partial\mathrm{Dom}(U)$. Applying the divergence theorem to $K_m$, we have

$$-\int_{K_m} x\cdot\nabla e^{-U(x)}dx = n\int_{K_m} e^{-U(x)}dx - \int_{K_m} e^{-U(x)}h_{K_m}(n_x)\ \mathrm{d}\sigma(x),$$

$$\leq n - \int_{\mathcal{B}_r(0)\bigcap\partial\mathrm{Dom}(U)} e^{-U(x)}h_{K_m}(n_x)\ \mathrm{d}\sigma(x).$$

Letting $m$ go to infinity, we arrive at a contradiction. Now since $\mathrm{int}(\mathrm{Dom}(U))$ is an open convex cone centered at the origin, we finally get that

$$U(tx) = tU(x) \quad \forall t > 0, \quad \forall x \in \mathrm{Int}(\mathrm{Dom}(U)).$$

- $4 \Rightarrow 2$.

  Define

$$C(p) := \frac{1}{\Gamma(n+p)}\int_{\mathbb{R}^n} U(x)^p e^{-U(x)}\ \mathrm{d}x$$

Since $\nu$ is log-concave of order $n$, we have

$$\frac{1}{\Gamma(n+p)}\int_{\mathbb{R}^n} U(x)^p e^{-U(x)}\ \mathrm{d}x = \frac{1}{\Gamma(n+p)}\int_0^\infty t^{n+p-1}g(t)\ \mathrm{d}t,$$

for some log-concave function $g(t)$ defined on $[0,\infty)$.
By Theorem 3.2.9, we have that $C(p)$ is log-concave on $[0,\infty)$. Therefore,

$$C(1)^2 \geq C(0)C(2),$$

which reads as

$$\int_{\mathbb{R}^n} U(x)^2 e^{-U(x)}dx \leq \frac{1+n}{n}\left(\int_{\mathbb{R}^n} U(x)e^{-U(x)}dx\right)^2.$$

We deduce further that

$$\mathrm{Var}(U(X)) \leq \frac{1}{n}\left(\mathbb{E}(U(X))\right)^2.$$

Since $\mathrm{Var}(U(X)) = n$ and $\mathbb{E}(U(X)) \leq n$, we have

$$\mathbb{E}(U(X)) = n,$$

as desired.

- $5 \Rightarrow 1$.

  By Lemma 3.2.6, equation (3.14), we have that $V(X_\alpha)$ is real analytic for $\alpha > 0$. Therefore $V(X_\alpha) = n$, for all $\alpha > 0$. Now use Lemma 3.2.8 to conclude.

- $6 \Rightarrow 1$.

  Recall the Definition of $F(\alpha)$, for each $\alpha > 0$,

$$F(\alpha) = \log \int f^\alpha(x) \, \mathrm{d}x.$$

  From the Taylor-Lagrange formula, for every $\alpha > 0$, we have

$$F(\alpha) = F(1) + F'(1)(\alpha - 1) + \int_1^\alpha F''(u)(\alpha - u)du.$$

  Note that $F(1) = 0$, and $F''(u) = \frac{V(X_u)}{u^2} \leq \frac{n}{u^2}$. Therefore, for $0 < \alpha < u < 1$, we have

$$F(\alpha) \leq F'(1)(\alpha - 1) + n \int_1^\alpha \frac{(\alpha - u)}{u^2} du. \tag{3.23}$$

  This inequality is obviously true for $\alpha \geq 1$. Thus, after a direct calculation, we have shown, for $\alpha > 0$,

$$F(\alpha) \leq F'(1)(\alpha - 1) + n(\alpha - 1 - \log \alpha).$$

  Set $\beta = 1 - \alpha$, and note that $e^{F(1-\beta)} = \mathbb{E}\left[e^{\beta U(X)}\right]$ and $e^{-\beta F'(1)} = e^{\beta \mathbb{E} U(X)}$, hence the inequality (3.23) can be written as

$$\mathbb{E} e^{\beta[U(X) - \mathbb{E}U(X)]} \leq e^{-n[\beta + \log(1-\beta)]}, \quad \forall \beta < 1.$$

  By continuity of $F(\alpha)$, the equality holds only if

$$F(\alpha) = \frac{n}{\alpha^2}, \quad \forall \alpha > 0,$$

  which implies that

$$V(X_\alpha) = n, \quad \forall \alpha > 0.$$

  We can now use Lemma 3.2.8 to complete the proof.

□

Though in general, it is difficult to check the log-concavity of the probability measure $\nu$, we manage to show an upper bound for the density function of $\nu$. This upper bound, once achieved, again implies the homogeneity of the potential function.

**Proposition 3.2.10.** *Consider a log-concave random vector $X$ in $\mathbb{R}^n$ with density function $f(x) = e^{-U(x)}$ such that $\min_{x\in\mathbb{R}^n} U(x) = U(0) = 0$. Let $g(t)$ be the density of the random variable $U(X)$. Then*

$$g(t) \leq nt^{n-1}e^{-t}\mathcal{L}^n(\{U(tx) \leq t\}), \quad \forall t > 0,$$

*where $\mathcal{L}^n$ represents the Lebesgue measure on $\mathbb{R}^n$. If the equality holds for all $t > 0$, then $U(x)$ is positively homogeneous of degree 1.*

*Proof.* Since $\mathrm{Dom}(U)$ is convex, its boundary has Lebesgue measure 0. Hence, $\mathbb{P}\{X \in \mathrm{int}(\mathrm{Dom}(U))\} = 1$. Therefore, we may restrict ourselves on the open convex set $\mathrm{int}(\mathrm{Dom}(U))$, where $\nabla U(x)$ is well-defined almost everywhere and locally bounded. By the Coarea formula, we have, for $0 \leq s < t$,

$$\int_{s<U(x)\leq t} e^{-U(x)}\,\mathrm{d}x = \int_s^t \int_{U(x)=r} \frac{e^{-U(x)}}{\|\nabla U(x)\|_2} dH^{n-1}\,\mathrm{d}r,$$

where $H^{n-1}$ is the $n-1$ dimensional Hausdorff measure. Taking derivative with respect to $t$, one gets the density for the probability measure $\nu$,

$$g(t) = e^{-t}\int_{U(x)=t} \frac{1}{\|\nabla U(x)\|_2} dH^{n-1} \quad a.e..$$

Note that $U(0) = 0$. By convexity of $U(x)$, for all points $x$ where $U(x)$ is differentiable, one has

$$U(x) \leq \langle \nabla U(x), x\rangle.$$

Thus, we have

$$e^{-t}\int_{U(x)=t} \frac{1}{\|\nabla U(x)\|_2} dH^{n-1} \leq \frac{e^{-t}}{t}\int_{U(x)=t} \frac{\langle \nabla U(x), x\rangle}{\|\nabla U(x)\|_2} dH^{n-1}.$$

By the divergence theorem, we have

$$g(t) \leq \frac{ne^{-t}}{t}\mathcal{L}^n(\{U(x) \leq t\}).$$

Use the homogeneity of the Lebesgue measure, we obtain

$$g(t) \leq nt^{n-1}e^{-t}\mathcal{L}^n(\{U(tx) \leq t\}),$$

as desired.

If the equality holds for all $t > 0$, then

$$h(X) = \int_0^\infty tg(t)\ \mathrm{d}t = n\int_0^\infty t^n e^{-t}\mathcal{L}^n(\{U(tx) \leq t\})\ \mathrm{d}t = n\int_{\mathbb{R}^n} e^{-U(x)}\ \mathrm{d}x = n.$$

The homogeneity now follows from Theorem 3.2.2.

□

### 3.3 Central limit theorem of the information content

In this section, we will show (see Theorem 3.3.6) that the bound in (3.2) is optimal in the sense that there exists a sequence of log-concave random vectors $X^{(n)}$ where, for each $n$, $X^{(n)}$ is in $\mathbb{R}^n$, such that for all $t > 0$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{\frac{\tilde{h}(X^{(n)}) - h(X^{(n)})}{n} > t\right\} = \log(1+t) - t. \tag{3.24}$$

Meanwhile, the optimal concentration inequalities (3.2) and (3.3) indicates that $-\log f(X)$ "behaves" like a sum of i.i.d. random variables. This leads us naturally to the following Central Limit Theorem (CLT) conjecture for the information content.

**Conjecture 8.** For sufficiently large $n$, given a log-concave random vector $X$ in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$, suppose that

$$V(X) = \beta n,\ \beta > 0.$$

Then,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{U(X)-\mathbb{E}[U(X)]}{\sqrt{V(X)}}\le t\right\}-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-\frac{r^2}{2}}\,\mathrm{d}r\right|\le\frac{c}{\sqrt{n}},$$

where $c$ is an absolute constant depending on $\beta$.

*Remark.* Note that $\beta \le 1$ by Theorem 3.2.1, and by no means is the condition $V(X) = \beta n$ optimal, but restriction on $V(X)$ is indeed required to make the conjecture meaningful. For example, consider the following log-concave densities

$$f(x) = e^{-x_1}\frac{\mathbf{1}_{K_{n-1}}(x_2,\cdots,x_n)}{|K_{n-1}|}, \quad x_1 > 0.$$

where $K_{n-1}$ is an convex body (compact convex set with nonempty interior) in $\mathbb{R}^{n-1}$ with Lebesgue measure 1, and $|K_{n-1}|$ denotes the Lebesgue measure of the set $K_{n-1}$ in $\mathbb{R}^{n-1}$. It is easy to check that $V(X) = 1$ and $U(X) - \mathbb{E}[U(X)] = X_1 - 1$, therefore, the conjecture is false.

Using the weak convergence of probability measures (random variables), one may also formulate a non-quantitative version of the Conjecture 8. In the sequel, let $\mathbb{X} = (X_1, X_2, \ldots)$ be a stochastic process on the probability space $(\Omega, \mathcal{B}, \mathbb{P})$, with each $X_i$ taking values in $\mathbb{R}$, and define the corresponding projections $X^{(n)} = (X_1, \ldots, X_n)$.

**Conjecture 9.** Suppose that $\mathbb{X}$ has a log-concave distribution on $\mathbb{R}^{\mathbb{N}}$ with absolutely continuous (with respect to Lebesgue measure) finite-dimensional projections, and for each $n$, denote by $f_n(x) = e^{-U_n(x)}$ the density of the projection $X^{(n)}$. Assume further that

$$\lim_{n\to\infty}\frac{V(X^{(n)})}{n} = \beta,\ \beta > 0.$$

Then,

$$\frac{U_n(X^{(n)}) - \mathbb{E}U_n(X^{(n)})}{\sqrt{V(X^{(n)})}} \rightharpoonup \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ is the standard normal random variable.

Conjecture 8 and Conjecture 9 follow from the classic central limit theorem for i.i.d. sequence of random variables when one considers a random vector $X$ in $\mathbb{R}^n$ or a stochastic process $\mathbb{X}$, with each coordinate drawn independently from a log-concave random variable $Y$ in $\mathbb{R}$ such that $V(Y) \in (0,1]$. Given that the classic central limit theorem can be generalized to independent but non-identically distributed random variables, we have the following proposition.

**Proposition 3.3.1.** *Let $\mathbb{X}$ be a stochastic process with independent coordinates, and for each $i$, the $i$-th coordinate $X_i$ is a log-concave random variable in $\mathbb{R}$ with log-concave density $g_i(x) = e^{-u_i(x)}$. Assume further that,*

$$\limsup_{n\to\infty} \frac{V(X^{(n)})}{n} > 0. \tag{3.25}$$

*Then,*

$$\frac{U_n(X^{(n)}) - \mathbb{E}U_n(X^{(n)})}{\sqrt{V(X^{(n)})}} \rightharpoonup \mathcal{N}(0,1), \tag{3.26}$$

*where $\mathcal{N}(0,1)$ is the standard normal random variable. Moreover, for sufficiently large $n$, we have the following quantitative result,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{U_n(X^{(n)}) - \mathbb{E}U_n(X^{(n)})}{\sqrt{V(X^{(n)})}} \le t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \,\mathrm{d}r \right| \le \frac{c}{\sqrt{n}}, \tag{3.27}$$

*where $c$ is an absolute constant independent of $\mathbb{X}$ and $n$.*

*Proof.* Since $\mathbb{X}$ has independent coordinates, for each $n$, the density of the projection $X^{(n)}$ is

$$f_n(x_1, \dots, x_n) = \prod_{i=1}^{n} g_i(x_i) = e^{-\sum_{i=1}^{n} u_i(x_i)}.$$

Therefore,

$$U_n(X^{(n)}) = \sum_{i=1}^{n} u_i(X_i).$$

Set $Y_i = u_i(X_i) - \mathbb{E}u_i(X_i)$ and let $S_n = \sum_{i=1}^n Y_i$. Define $s_n^2 := \mathrm{Var}(S_n) = V(X^{(n)})$. We want to show that

$$\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}} \rightharpoonup \mathcal{N}(0,1).$$

To this end, for a given $\epsilon > 0$, by Fubini theorem, we have

$$\mathbb{E}[Y_i^2 \mathbf{1}_{\{|Y_i| \geq \epsilon s_n\}}] = \int_\Omega \mathbf{1}_{\{|Y_i| \geq \epsilon s_n\}} \int_0^{|Y_i|} 2t \ \mathrm{d}t \ \mathrm{d}\mathbb{P}$$

$$= 2\int_0^\infty t \ \mathbb{P}\{|Y_i| \geq \max\{\epsilon s_n, t\}\} dt = (\epsilon s_n)^2 \mathbb{P}\{|Y_i| \geq \epsilon s_n\} + 2\int_{\epsilon s_n}^\infty t \ \mathbb{P}\{|Y_i| \geq t\} \ \mathrm{d}t. \tag{3.28}$$

By inequalities (3.2) and (3.3), for $n$ large enough such that $s_n > \frac{1}{\epsilon}$, we have the following estimation

$$(\epsilon s_n)^2 \mathbb{P}\{|Y_i| \geq \epsilon s_n\} + 2\int_{\epsilon s_n}^\infty t \ \mathbb{P}\{|Y_i| \geq t\} \ \mathrm{d}t$$

$$\leq (\epsilon s_n)^2 (1 + \epsilon s_n) e^{-\epsilon s_n} + 2\int_{\epsilon s_n}^\infty t(1+t)e^{-t} \ \mathrm{d}t \leq c(\epsilon)(s_n)^3 e^{-\epsilon s_n}, \tag{3.29}$$

where $c(\epsilon)$ is an absolute constant that only depends on $\epsilon$.

Combining equation (3.28) and inequality (3.29), we have

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}[Y_i^2 \mathbf{1}_{\{|Y_i| \geq \epsilon s_n\}}] \leq c(\epsilon) n s_n e^{-\epsilon s_n}. \tag{3.30}$$

Our assumption implies that

$$n s_n e^{-\epsilon s_n} \to 0,$$

as $n \longrightarrow \infty$. By Lindeberg-Feller theorem (see, e.g., [38]), the qualitative result (3.26) follows.

To prove inequality (3.27), use the layer cake representation and inequalities (3.2) and

(3.3), once again, we have

$$\mathbb{E}|Y_i|^3 = 3\int_0^\infty t^2\ \mathbb{P}\{|Y_i| \geq t\}\}dt \leq 3 + 3\int_1^\infty t^2\ \mathbb{P}\{|Y_i| \geq t\}\ \mathrm{d}t \leq c,$$

where $c$ is an absolute constant. Combining our assumption (3.25) with the Berry-Esseen theorem for non-identically distributed random variables (see, e.g., [54]), the desired quantitative result follows. □

*Remark.* Inequality (3.30) implies that the qualitative result (3.26) actually holds under a weaker condition

$$\limsup_{n\to\infty} \frac{V(X^{(n)})}{n^\delta} > 0. \tag{3.31}$$

for any fixed $\delta > 0$.

However, our following example shows that condition (3.31) is far from optimal, in the sense that there exists a stochastic process $\mathbb{X}$ satisfying a weaker condition, while the qualitative result (3.26) still holds.

**Example 1.** Let $\mathbb{X}$ be a stochastic process with independent coordinates, and for each $i$, the $i$-th coordinate $X_i$ is a log-concave random variable in $\mathbb{R}$ with log-concave density $g_i(x) = e^{-c_i|x|^i}$, where $c_i$ is a normalization constant.

To see that the example indeed satisfies a weaker condition, note that $U_n(X^{(n)}) = \sum_{i=1}^n c_i|X_i|^i$, and a direct calculation reveals that

$$V(X^{(n)}) = \mathrm{Var}(U_n(X^{(n)})) = \sum_{i=1}^n \frac{1}{i} \sim \log n.$$

To show the result (3.26), recall the Lyapunov's theorem (see, e.g., [38]):

**Theorem 3.3.2.** *Let $Y_1, Y_2, \cdots$ be independent random variables, and $S_n = \sum_{i=1}^n Y_i$. Let $\alpha_n = \sqrt{Var(S_n)}$, if there exists a $\delta > 0$, such that*

$$\lim_{n\to\infty} \alpha_n^{-(2+\delta)} \sum_{i=1}^n \mathbb{E}(|Y_i - \mathbb{E}Y_i|^{2+\delta}) = 0,$$

*then we have*

$$\frac{S_n - \mathbb{E}S_n}{\alpha_n} \rightharpoonup \mathcal{N}(0, 1).$$

For our example, set $Y_i = c_i|X_i|^i$, and choose $\delta = 2$. A direct computation tells us that

$$\sum_{i=1}^n \mathbb{E}(|Y_i - \mathbb{E}Y_i|^4) = 3\sum_{i=1}^n (\frac{1}{i^2} + \frac{2}{i}).$$

By Lyapunov's theorem, the desired result follows.

It is unclear to us yet what the minimum threshold that $V(X^{(n)})$ needs to exceed to guarantee the qualitative central limit theorem. However, if we content ourselves with the condition assumed in Conjecture 9, Proposition 3.3.1 provides an affirmative answer to the Conjecture 9 when we also assume independence of the coordinates. And we will show shortly that by trading independence for uniform boundedness of a higher moment, one can give a unified approach to Conjecture 9. The following result is credited to L. Wang [103], we give a proof here for completeness.

For a log-concave random vector $X$ with density $f(x) = e^{-U(x)}$, recall that $K_3(X)$ represents the third moment of the random variable $U(X) - \mathbb{E}U(X)$, and $X_\alpha$ is a random vector with density $f_\alpha = \frac{f^\alpha}{\int_{\mathbb{R}^n} f^\alpha \, \mathrm{d}x}$.

**Theorem 3.3.3.** *Under the assumptions in Conjecture 9, suppose, in addition, that there exists an absolute constant $\delta > 0$ such that*

$$\lim_{n\to\infty} \sup_{\alpha\in(1-\delta,1+\delta)} \frac{|K_3(X_\alpha^{(n)})|}{n^{\frac{3}{2}}} = 0.$$

*Then Conjecture 9 holds.*

*Proof.* Define $F_n(\alpha) = \log \int_{\mathbb{R}^n} f_n^{\alpha}(x)\,\mathrm{d}x$. By Lemma 3.2.6, $F_n(\alpha)$ is real analytic on $\alpha > 0$, and is therefore infinitely differentiable. A direct calculation reveals that:

$$F_n'(\alpha) = \int_{\mathbb{R}^n} f_{n,\alpha} \log f_n(x)\,\mathrm{d}x,$$

where $f_{n,\alpha} = \frac{f_n^{\alpha}}{\int_{\mathbb{R}^n} f_n^{\alpha}\,\mathrm{d}x}$.*Similarly,*

$$F_n''(\alpha) = \frac{V(X_\alpha^{(n)})}{\alpha^2},$$

$$F_n^{(3)}(\alpha) = -\frac{K_3(X_\alpha^{(n)})}{\alpha^3},$$

where $X_\alpha^{(n)}$ is the random vector with density $f_{n,\alpha}$.

For every $\alpha > 0$, by Taylor's theorem, we have

$$F_n(\alpha) = F_n(1) + F_n'(1)(\alpha-1) + F_n''(1)\frac{(\alpha-1)^2}{2!} + \frac{1}{2!}\int_1^{\alpha} F_n^{(3)}(u)(\alpha-u)^2 du.$$

Setting $\alpha = 1$, we obtain $F_n(1) = 0, F_n'(1) = -\mathbb{E}[U_n(X^{(n)})]$, and $F_n''(1) = V(X^{(n)})$. Therefore,

$$F_n(\alpha) = -\mathbb{E}[U_n(X^{(n)})](\alpha-1) + V(X^{(n)})\frac{(\alpha-1)^2}{2!} - \frac{1}{2!}\int_1^{\alpha} \frac{K_3(X_u^{(n)})}{u^3}(\alpha-u)^2 du.$$

Observe that,

$$e^{F_n(\alpha)} = \int f_n^{\alpha} = \mathbb{E}e^{(\alpha-1)\log f_n(X^{(n)})} = \mathbb{E}e^{-(\alpha-1)U_n(X^{(n)})}.$$

Hence,

$$\log \mathbb{E}e^{-(\alpha-1)\{U_n(X^{(n)})-\mathbb{E}[U_n(X^{(n)})]\}} = V(X^{(n)})\frac{(\alpha-1)^2}{2!} - \frac{1}{2!}\int_1^{\alpha} \frac{K_3(X_u^{(n)})}{u^3}(\alpha-u)^2 du. \tag{3.32}$$

Note that

$$\left|\int_1^{\alpha} \frac{K_3(X_u^{(n)})(\alpha-u)^2}{u^3}du\right| \leq \sup_{u\in[1,\alpha]} \left|K_3(X_u^{(n)})\right| \left(\log\alpha - 2\alpha + \frac{\alpha^2}{2} + 3/2\right). \tag{3.33}$$

For $t>0$, set $\alpha = 1 + \frac{t}{\sqrt{V(X^{(n)})}}$ in equation (3.32). By inequality (3.33), we have

$$\left|\log \mathbb{E}e^{-t\left(\frac{U_n(X^{(n)})-\mathbb{E}[U_n(X^{(n)})]}{\sqrt{V(X^{(n)})}}\right)} - \frac{t^2}{2}\right|$$

$$\leq \frac{\sup_{u\in[1,\alpha]}\left|K_3(X_u^{(n)})\right|}{2}\left(\log\left(1+\frac{t}{\sqrt{V(X^{(n)})}}\right) - \frac{t}{\sqrt{V(X^{(n)})}} + \frac{1}{2}\frac{t^2}{V(X^{(n)})}\right). \tag{3.34}$$

For $x>0$ small enough, we have

$$\log(1+x) - x + \frac{x^2}{2} = \frac{x^3}{3} + o(x^3). \tag{3.35}$$

For $n$ large enough, by the assumption, one has

$$\frac{\sup_{u\in[1,\alpha]}\left|K_3(X_u^{(n)})\right|}{V(X^{(n)})^{3/2}} = \frac{\sup_{u\in[1,\alpha]}\left|K_3(X_u^{(n)})\right|}{n^{3/2}}\frac{n^{3/2}}{V(X^{(n)})^{3/2}} \longrightarrow 0. \tag{3.36}$$

Combining (3.34), (3.35), and (3.36), we obtain

$$\lim_{n\to\infty} \mathbb{E}e^{-t\left(\frac{U_n(X^{(n)})-\mathbb{E}[U_n(X^{(n)})]}{\sqrt{V(X^{(n)})}}\right)} = e^{\frac{t^2}{2}}, \ \forall t>0.$$

That is, the Laplace transform of $\frac{U_n(X^{(n)})-\mathbb{E}[U_n(X^{(n)})]}{\sqrt{V(X^{(n)})}}$ converges to the Laplace transform of $\mathcal{N}(0,1)$. The desired result now follows.

□

As an immediate consequence of Theorem 3.3.3, Theorem 3.2.2, and Proposition 3.2.7 we have the following corollary.

**Corollary 3.3.4.** *Suppose that* $\mathbb{X}$ *has a log-concave distribution on* $\mathbb{R}^{\mathbb{N}}$ *with absolutely continuous (with respect to Lebesgue measure) finite-dimensional projections, and for each* $n$*, denote by* $f_n(x) = e^{-U_n(x)}$ *the density of the projection* $X^{(n)}$*. Assume further that, for each* $n$*,* $\sup_{x\in\mathbb{R}^n} f_n(x) = 1$ *and*

$$h(X^{(n)}) = n.$$

*Then,*

$$\frac{U_n(X^{(n)}) - \mathbb{E}U_n(X^{(n)})}{\sqrt{V(X^{(n)})}} \rightharpoonup \mathcal{N}(0,1),$$

*where* $\mathcal{N}(0,1)$ *is the standard normal random variable.*

In fact, the assumption in Corollary 3.3.4 enables us to get quantitative results, since in Theorem 3.2.2, we have shown that $U(X) \stackrel{d}{=} \text{Gamma}(n,1)$. But in order to state our result in a systematic way, we need some preliminaries. The following definition can be found in (see, e.g., [82]).

**Definition 11.** A function $f$ defined on $(0,\infty)$ is completely monotone if it has derivatives of all orders and

$$(-1)^k f^{(k)}(t) > 0, \quad t \in (0,\infty),\ k = 0,1,2,\ldots$$

A function $f(x)$ defined on $(0,\infty)$ is said to have a completely monotone derivative if $f'(x)$ is completely monotone.

**Definition 12.** A measure $\mu$ on $\mathbb{R}^d$ is called infinitely divisible (ID) if for every natural number $n$, there exists a measure $\mu_n$ on $\mathbb{R}^d$ such that

$$\mu = \underbrace{\mu_n * \mu_n * \ldots * \mu_n}_{n},$$

where $*$ represents convolution.

Thanks to W. Feller [40], one may set up a connection between completely monotone functions and infinitely divisible probability measures.

**Theorem 3.3.5.** *A function $f$ is the Laplace transform of an infinitely divisible probability measure if and only if*

$$f(x) = e^{-g(x)},$$

*where $g(x)$ has a completely monotone derivative and $g(0) = 0$.*

We are now ready to state our main theorem in this section

**Theorem 3.3.6.** *Given a log-concave random vector $X$ in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$, suppose that $U(x)$ is positively homogeneous of degree $\gamma \geq 1$, that is*

$$U(tx) = t^\gamma U(x), \quad t > 0, \quad x \in \mathbb{R}^n.$$

*Then,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{U(X) - \mathbb{E}U(X)}{\sqrt{V(X)}} \leq t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}} \,\mathrm{d}r \right| \leq \frac{c}{\sqrt{n}},$$

*where $c$ is an absolute constant that only depends on $\gamma$.*

*Moreover, if $X^{(n)}$ is a sequence of log-concave random vectors such that, for each $n$, $X^{(n)}$ is in $\mathbb{R}^n$ with density $f_n(x) = e^{-U_n(x)}$, and $U_n(tx) = t^\gamma U_n(x)$, $\forall t > 0$, $\forall x \in \mathbb{R}^n$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{ \frac{U_n(X^{(n)}) - \mathbb{E}U_n(X^{(n)})}{V(X^{(n)})} > t \right\} = \frac{\log(1+t) - t}{\gamma}.$$

*Proof.* For any $t > 0$, the Laplace transform of the random variable $U(X)$ is

$$\mathbb{E}e^{-tU(X)} = \int_{\mathbb{R}^n} e^{-(1+t)U(x)} \,\mathrm{d}x.$$

Using the homogeneity of $U(x)$, we have

$$\int_{\mathbb{R}^n} e^{-(1+t)U(x)} \,\mathrm{d}x = (1+t)^{-\frac{n}{\gamma}}.$$

Note that the function

$$g(t) = \frac{n}{\gamma}\log(1+t)$$

has a completely monotone derivative, and $g(0) = 0$. Therefore, the law of $U(X)$ is infinitely divisible by Theorem 3.3.5. Now from the Laplace transform of $U(X)$, one can easily deduce that

$$U(X) \overset{d}{=} \sum_{i=1}^{n} Y_i,$$

where, for each $i$, $Y_i$ is an independent copy of the gamma distribution with shape parameter $\frac{1}{\gamma}$ and scale parameter 1. The desired results follow from the Berry-Esseen theorem and the large deviation principle result for i.i.d. gamma distributions.

□

Combining our Theorem 3.3.6 with Theorem 3.2.2, we obtain the following corollary.

**Corollary 3.3.7.** *Given a log-concave random vector $X$ in $\mathbb{R}^n$ with density $f(x) = e^{-U(x)}$, suppose that $min_{x\in\mathbb{R}^n} U(x) = U(0) = 0$, and*

$$h(X) = n.$$

*Then,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left\{ \frac{U(X) - \mathbb{E}U(X)}{\sqrt{V(X)}} \le t \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{r^2}{2}}\, \mathrm{d}r \right| \le \frac{c}{\sqrt{n}},$$

*where $c$ is an absolute constant.*

*Proof.* By our assumptions, Theorem 3.2.2 implies that $U(x)$ is necessarily a positively homogeneous function of degree 1. The desired result follows by Theorem 3.3.6.

□

Though our main results deal with homogeneous potentials, the next example shows that Conjecture 8 can still hold with non-homogeneous potentials.

**Example 2.** Let $E$ be a $k$ dimensional subspace of $\mathbb{R}^n$, and $P$ be the coordinate (orthogonal) projection from $\mathbb{R}^n$ to $E$. Let $K$ and $L$ be two origin symmetric convex bodies living in $E$ and $E^\perp$, respectively. Consider the convex function $U(x) = c_1\|Px\|_K + c_2\|(I_n - P)x\|_L^2$, where $\|\cdot\|_K$ and $\|\cdot\|_L$ are norms induced by the convex bodies $K$ and $L$, respectively, and the positive constants $c_1$ and $c_2$ are chosen such that $f(x) = e^{-U(x)}$ is a probability density. Let $X$ be a log-concave random vector in $\mathbb{R}^n$ with density function $f(x)$, then the Conjecture 8 holds.

Indeed, one has

$$\int_{U(x)\leq t} e^{-U(x)}dx = \int_{c_1\|Px\|_K + c_2\|(I-P)x\|_L^2 \leq t} e^{-c_1\|Px\|_K} e^{-c_2\|(I-P)x\|_L^2}\, \mathrm{d}x.$$

Therefore, by Theorem 3.3.6, we have

$$U(X) \stackrel{d}{=} \mathrm{Gamma}(k,1) + \mathrm{Gamma}(\frac{n-k}{2}, 1).$$

As a consequence, $V(X) = \mathrm{Var}(U(X)) = \frac{n+k}{2}$, and $\frac{V(X)}{n} \in [1/2, 1]$. Conjecture 8 follows from the Berry-Esseen theorem for non-identically distributed random variables.

Without too much effort, we can generalize Example 2 to the following theorem.

**Theorem 3.3.8.** *Let $\beta > 0$ be a constant and suppose that $\mathbb{R}^n = E_1 \oplus E_2 \oplus \ldots \oplus E_m$ with dim$(E_i) = k_i$. Let $\gamma = (\gamma_1, \ldots, \gamma_m) \in [1, \infty)^m$ such that*

$$\sum_{i=1}^{m} \frac{k_i}{\gamma_i} = \beta n.$$

*Also, let $\|\cdot\|_{E_i}$ be any norm in each $E_i$, select $c = (c_1, \cdots, c_m) \in \mathbb{R}_+^m$, such that*

$$f(x) = e^{-\sum_{i=1}^m c_i \|P_{E_i}x\|_{E_i}^{\gamma_i}},$$

*is a probability density. Let $X$ be a log-concave random vector in $\mathbb{R}^n$ with density $f(x)$. Then Conjecture 8 holds.*

*Proof.* Since the quantity

$$\frac{U(X) - \mathbb{E}U(X)}{\sqrt{V(X)}}$$

is affine invariant, hitting $X$ with a suitable orthogonal transformation, if necessary, we may assume that $P_{E_i}$ are coordinate projections. By Theorem 3.3.6, we have

$$U(X) \stackrel{d}{=} \sum_{i=1}^{m} \text{Gamma}(\frac{k_i}{\gamma_i}, 1) \stackrel{d}{=} \text{Gamma}(\beta n, 1).$$

A direct calculation reveals that

$$V(X) = \sum_{i=1}^{m} \frac{k_i}{\gamma_i}.$$

The desired result now follows from our assumption and the Berry-Esseen theorem.

□

# Chapter 4

# DIVERSITY AND DIMENSION OF PROBABILITY MEASURES IN METRIC SPACES

## 4.1 Diversity and dimension of order $\alpha$

As discussed in the introduction, we model the setup of a population and the associated measure of similarity by a compact Hausdorff space $X$ and a similarity kernel $K : X \times X \to [0, \infty)$, respectively. Here we think of $K(x, y)$ as a measure of how similar the point $x$ is to the point $y$. While most naturally occuring notions of similarity are symmetric, *a priori* we do not impose this restriction on our similarity kernels.

**Definition 13.** A **similarity kernel** on a Hausdorff topological space $X$ is a continuous function $K : X \times X \to [0, \infty)$ such that $K(x, x) > 0$ for every $x \in X$. The pair $(X, K)$ will be referred to as a **space with similarities**.

**Remark 4.1.1.** *Leinster and Roff [66] define the similarity kernel for compact Hausdorff spaces but we are dropping the compactness assumption here. For dealing with maximum diversity though, we will require $X$ to be compact as in [66].*

**Examples 4.1.2.**

1. *The* ***Dirac kernel*** *on finite spaces equipped with discrete topology,*

$$K(x, y) = \delta_{x,y} = \begin{cases} 0, & \text{if } x \neq y \\ 1, & \text{if } x = y, \end{cases}$$

*is continuous. Note that finiteness of $X$ is required for $K$ to be continuous. However, one can approximate the (discontinuous) Dirac kernel on a (possibly infinite) compact metric space by a family of appropriately scaled Laplace kernels (see below). This approximation will in fact play an important role in our analysis later.*

2. *The **Laplace kernel** on a metric space $(X, d)$ is defined as*

$$K(x, y) = e^{-d(x,y)},$$

*for $x, y \in X$. Moreover, $(X, td)$, for $t > 0$, is also a metric space. The corresponding Laplace kernel will be denoted by $K^t(x, y) = e^{-td(x,y)}$. Note that the Laplace kernel tensorizes over the $\ell_1$-product of metric spaces. More precisely, if $(X_1, d_1), (X_2, d_2)$ are metric spaces with corresponding Laplace kernels $K_1, K_2$ respectively, then, the Laplace kernel on the metric space $X_1 \times_1 X_2$ is given by*

$$K((x_1, x_2), (y_1, y_2)) = K_1 \otimes K_2((x_1, x_2), (y_1, y_2))$$
$$= e^{-d((x_1,x_2),(y_1,y_2))} = e^{-(d_1(x_1,y_1)+d_2(x_2,y_2))}.$$

*Here $X_1 \times_1 X_2$ is the cartesian product $X_1 \times X_2$ equipped with the metric*

$$d((x_1, x_2), (y_1, y_2)) = d_1(x_1, y_1) + d_2(x_2, y_2).$$

3. *A very natural kernel to consider from a probabilist's perspective, at least in Euclidean space, is the **Gaussian kernel***

$$K(x, y) = e^{-d(x,y)^2},$$

*for $x, y \in X$. Just as the Laplace kernel tensorizes over the $\ell_1$-product of metric spaces, the Gausssian kernel tensorizes over the $\ell_2$-product, $X_1 \times_2 X_2$, of metric spaces $(X_1, d_1)$ and $(X_2, d_2)$, which is the cartesian product $X_1 \times X_2$ equipped with the metric $d((x_1, x_2), (y_1, y_2)) = \left(d_1(x_1, y_1)^2 + d_2(x_2, y_2)^2\right)^{1/2}$. In the Euclidean case, as we will later see, the Gaussian kernel lets us exploit tools from mass transport to obtain sumset inequalities for Minkowski dimension.*

4. *Reflexive binary relations on finite sets give rise to kernels in the following natural manner. Suppose $X$ is a finite set (which we think of as a discrete topological space) and $R \subseteq X \times X$ a reflective binary relation. Then one can define the binary-valued **adjacency kernel** by*

$$K_R(x, y) = \begin{cases} 1, & \text{if } (x, y) \in R \\ 0, & \text{otherwise.} \end{cases}$$

*From here, one can obtain interesting sub-examples taking $R$ to be the adjacency in a graph, or the order in a partially-ordered set.*

We will denote by $\mathcal{P}(X)$ the space of Borel probability measures on the Hausdorff space $X$. A similarity kernel $K$ on $X$ naturally induces a map from the $\mathcal{P}(X)$ to the set of extended real-valued functions on $X$, via

$$K\mu(x) = \int K(x, y) \, \mathrm{d}\mu(y).$$

Therefore, $K\mu(x)$ measures how typical the point $x$ is, among all points in $X$, when sampled according to $\mu$. Of course, $1/K\mu(x)$ then measures the *atypicality* of $x \in X$. Various measures of average atypicality yield the notions of entropy that are at the center of our investigations.

Let $(X, K)$ be a space with similarities and $\mu \in \mathcal{P}(X)$. Recall the **diversity of order** $\alpha$ of $\mu$ with respect to the kernel $K$ defined in Definition 8.

$$D_\alpha^K(\mu) = \begin{cases} \left(\int (1/K\mu(x))^{1-\alpha} \, \mathrm{d}\mu(x)\right)^{1/(1-\alpha)} & \text{if } \alpha \in [0,1) \cup (1, \infty), \\ e^{-\int \log K\mu(x) \, \mathrm{d}\mu(x)} & \text{if } \alpha = 1, \\ \frac{1}{\text{ess sup}_\mu K\mu(x)} & \text{if } \alpha = \infty, \end{cases}$$

where $K\mu(x) = \int K(x, y) \, \mathrm{d}\mu(y)$, as defined earlier.

**Remark 4.1.3.** *Diversities are the exponentials of the entropic notions. The corresponding entropy of order $\alpha$ of a probability measure $\mu$ on the space with similarities $(X, K)$ is therefore defined by,*

$$H_\alpha^K(\mu) = \log D_\alpha^K(\mu).$$

In Section 3, when we discuss maximum diversity we will restrict the discourse to compact spaces. Under this restriction, the typicality function $\mu \to K\mu$ defined $\mathcal{P}(X) \to C(X)$ is continuous [66]. As a consequence, diversities enjoy many more continuity properties on compact spaces, some of which we summarize now.

**Proposition 4.1.4.** *[66] Let $(X, K)$ be a space with similarities. Assume that the Hausdorff topological space $X$ is also compact. Then the real-valued function $D(\mu, \alpha) =$*

$D_\alpha^K(\mu)$ *defined on the product space* $\mathcal{P}(X) \times [0, \infty)$*, where* $\mathcal{P}(X)$ *is equipped with the topology of weak convergence of measures, is coordinate-wise continuous.*

We now proceed to make an explicit connection between the diversity measures introduced above and their classical information-theoretic counterparts, namely Rényi entropies. Recall the definition of Rényi entropy of order $\alpha$: For a $\mathbb{R}^n$-valued random variable with probability density $f$, its Rényi entropy of order $\alpha$ is defined by

$$h_\alpha(f) := \frac{1}{1-\alpha} \log \int_{\mathbb{R}^n} f^\alpha(x) dx,$$

if $\alpha \in [0, 1) \cup (1, \infty)$. The Rényi entropies of order $\alpha = 1, \infty$ are obtained by taking limits.

Let us consider the special case when $X = \mathbb{R}^n$ and the similarity kernel $K$ is of the form $K(x, y) = k(\|x - y\|)$, where $k$ is a non-negative decreasing function and $k \circ \| \cdot \|$ is a probability density (this can always be achieved by normalization as long as $k \circ \| \cdot \|$ is integrable). Here $\| \cdot \|$ denotes any fixed norm on $\mathbb{R}^n$. In this scenario, $t^n K^t(0, \cdot) t^n := t^n k \circ t \| \cdot \|$ also becomes a probability density on $\mathbb{R}^n$ which weakly converges to $\delta_0$, the dirac mass at 0. Suppose the measure $\mu$ who diversities we wish to evaluate also has density $f$. Then the typicalities $(t^n K^t \mu)$ actually converge to the function $f$ pointwise as $t \to \infty$, yielding exponentials of Rényi entropies from $\alpha$-diversities in the limit. This is how the promised recovery of Rényi entropic measures is obtained. The details, more formally, are given below.

**Definition 14.** A similarity kernel $K$ on $\mathbb{R}^n$ is said to be contoured if it is generated by a decreasing continuous function $k(s) > 0$ on $[0, \infty)$, such that $K(x, y) = k(\|x - y\|)$, where $\| \cdot \|$ is a norm on $\mathbb{R}^n$. For $t > 0$, we can define a family of kernels as follows

$$K_t(x, y) := k(t\|x - y\|) \quad \forall x, y \in \mathbb{R}^n.$$

**Proposition 4.1.5** (Recovery of Renyi entropies in $\mathbb{R}^n$)**.** *Suppose the similarity kernel* $K$ *on* $\mathbb{R}^n$ *is contoured and generated by* $k(s)$*. Given* $\mu \in \mathcal{P}(\mathbb{R}^n)$ *with density* $f(x)$*, if*

*$k(\|x\|) \in L^1(\mathbb{R}^n)$, then for $\alpha \in (1,\infty)$,*

$$e^{h_\alpha(f)} = \lim_{t\to\infty} \left(\int_{\mathbb{R}^n} k(\|x\|)dx\right) \frac{D_\alpha^{K_t}(\mu)}{t^n}$$

*Proof.* By definition

$$\left(\int_{\mathbb{R}^n} k(\|x\|)dx\right) \frac{D_\alpha^{K_t}(\mu)}{t^n} = \left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \frac{t^n K_t(x,y)}{\int_{\mathbb{R}^n} k(\|x\|)dx} f(y)dy\right)^{\alpha-1} f(x)dx\right)^{\frac{1}{1-\alpha}}.$$

Therefore, without loss of generality, we can assume $\int_{\mathbb{R}^n} k(\|x\|)dx = 1$. Since $f \in L^1(\mathbb{R}^n)$, by [50, Corollary 2.1.17], we have

$$\int_{\mathbb{R}^n} t^n K_t(x,y) f(y)dy \to f(x) \quad a.e.$$

For $\alpha \in (1,\infty)$, by Fatou's lemma, we have

$$\liminf_{t\to\infty} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} t^n K_t(x,y) f(y)dy\right)^{\alpha-1} f(x)dx \geq \int_{\mathbb{R}^n} f^\alpha(x)dx.$$

If $\int_{\mathbb{R}^n} f^\alpha(x)dx = +\infty$, then trivially, we have

$$e^{h_\alpha(f)} = \lim_{t\to\infty} \left(\int_{\mathbb{R}^n} k(\|x\|)dx\right) \frac{D_\alpha^{K_t}(\mu)}{t^n} = 0.$$

If $\int_{\mathbb{R}^n} f^\alpha(x)dx < +\infty$, by [50, Theorem 1.2.19], we have

$$\int_{\mathbb{R}^n} t^n K_t(x,y) f(y)dy \to f(x) \quad \text{in} \quad L^\alpha(\mathbb{R}^n).$$

Meanwhile, note that $\alpha > 1$, by Holder's inequality,

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} t^n K_t(x,y) f(y)dy\right)^{\alpha-1} f(x)dx$$

$$\leq \left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} t^n K_t(x,y) f(y)dy\right)^{\alpha} dx\right)^{\frac{\alpha-1}{\alpha}} \left(\int_{\mathbb{R}^n} f^\alpha(x)dx\right)^{\frac{1}{\alpha}}.$$

Whence, we have

$$\limsup_{t\to\infty} \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} t^n K_t(x,y) f(y) dy \right)^{\alpha-1} f(x) dx \leq \int_{\mathbb{R}^n} f^\alpha(x) dx$$

and the result follows. □

As a corollary, we observe that the $\alpha$-diversities of probability measures having density with respect to the $n$-dimensional Lebesgue measure, under contoured kernels, scale asymptotically like $t^n$. Therefore, diversity measures detect the dimension of probability measures if they are absolutely continuous. This idea can be extended in general by defining the $\alpha$-dimension of a probability measure on a metric space as the asymptotic order of growth of the $\alpha$-diversity with respect to scaling of the associated Laplace kernel.

**Definition 15** ($\alpha$-dimension)**.** Let $(X, d)$ be a metric space, $K(x,y) = e^{-d(x,y)}$ the corresponding Laplace kernel, $\alpha \in (0, \infty)$. We define the upper and lower $\alpha$-dimensions of the measure $\mu \in \mathcal{P}(X)$ by

$$\overline{\dim}^\alpha(\mu) = \limsup_{t\to\infty} \frac{\log D_\alpha^{K^t}(\mu)}{\log t},$$

$$\underline{\dim}^\alpha(\mu) = \liminf_{t\to\infty} \frac{\log D_\alpha^{K^t}(\mu)}{\log t},$$

respectively. When $\underline{\dim}^\alpha(\mu) = \overline{\dim}^\alpha(\mu)$, we call this value the diversity dimension of order $\alpha$ of $\mu$ and denote it $\dim^\alpha(\mu)$.

We will show that this notion of dimension specializes to a well known notion of dimension, namely information dimension (defined below), when the underlying space is $\ell_2^n$.

**Definition 16** (Information dimension)**.** For $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, *let* $\lfloor x \rfloor$ denote the vector $(\lfloor x_1 \rfloor, \cdots, \lfloor x_n \rfloor)$. The upper and lower information dimensions of order

$\alpha \in (0,\infty)$ of a random vector $X$ in $\mathbb{R}^n$ are defined via the following limits:

$$\overline{d}^{\alpha}(X) = \limsup_{k\to\infty} \frac{H_\alpha(X_k)}{\log k}, \qquad \underline{d}^{\alpha}(X) := \liminf_{k\to\infty} \frac{H_\alpha(X_k)}{\log k},$$

where $H_\alpha(X_k)$ is the Rényi entropy of order $\alpha$ of the random vector $X_k \triangleq \frac{\lfloor kX \rfloor}{k}$.When $\underline{d}^{\alpha}(X) = \overline{d}^{\alpha}(X)$, we call this value the information dimension of order $\alpha$ of $X$, denoted $d^\alpha(X)$.

The following theorem reveals the connection between these two definitions of dimensions.

**Theorem 4.1.6** (diversity dim is info dim)**.** *Denote by $\mu$ the law of a random vector $X$. Then for $\alpha > 1$, $\overline{\dim}^{\alpha}(\mu) = \overline{d}^{\alpha}(X)$, and $\underline{\dim}^{\alpha}(\mu) = \underline{d}^{\alpha}(X)$. In particular,*

$$\dim^\alpha(\mu) = d^\alpha(X),$$

*if either of them exists. For $\alpha \in (0,1)$, if in addition, $\mu$ is compactly supported, then $\overline{\dim}^{\alpha}(\mu) = \overline{d}^{\alpha}(X)$, and $\underline{\dim}^{\alpha}(\mu) = \underline{d}^{\alpha}(X)$. In particular,*

$$\dim^\alpha(\mu) = d^\alpha(X),$$

*if either of them exists.*

*Proof.* For $k \geq 1$, set $A_{i_1,\cdots,i_n,k} = [i_1/k,(i_1+1)/k) \times \cdots \times [i_n/k,(i_n+1)/k)$ for each integer $i_m$, $m = 1,\cdots,n$, and let $p_{i_1,\cdots,i_n,k} = \mu(A_{i_1,\cdots,i_n,k})$. In the sequel, we use $\sum_i$ to represent the $n$-fold sum $\sum_{i_1\in\mathbb{Z}} \cdots \sum_{i_n\in\mathbb{Z}}$, and for a positive integer $l$, the $n$-fold sum $\sum_{i_1=-l}^{l} \cdots \sum_{i_n=-l}^{l}$ will be compactly written as $\sum_{i=-l}^{l}$. For $t > 0$, one has

$$\int \left( \int e^{-td(x,y)} \,\mathrm{d}\mu \right)^{\alpha-1} \mathrm{d}\mu = \sum_j \int_{A_{j_1,\cdots,j_n,k}} \left( \sum_i \int_{A_{i_1,\cdots,i_n,k}} e^{-td(x,y)} \,\mathrm{d}\mu \right)^{\alpha-1} \mathrm{d}\mu.$$

$$\geq \sum_j \int_{A_{j_1,\cdots,j_n,k}} \left( \int_{A_{j_1,\cdots,j_n,k}} e^{-td(x,y)} \,\mathrm{d}\mu \right)^{\alpha-1} \mathrm{d}\mu \geq \sum_j p^{\alpha}_{j_1,\cdots,j_n,k} e^{-\sqrt{n}(\alpha-1)t/k}.$$

Therefore

$$\frac{1}{1-\alpha}\log\int\left(\int e^{-td(x,y)}\ \mathrm{d}\mu\right)^{\alpha-1}\ \mathrm{d}\mu \le H_\alpha(X_k)+\sqrt{n}t/k.$$

Setting $t=k$ and letting $k\to\infty$, one obtains

$$\underline{\dim}^\alpha(\mu)\le\liminf_{k\to\infty}\frac{\log D_\alpha^{K^n}(\mu)}{\log k}\le\underline{d}^\alpha(X).$$

Meanwhile, setting $k=\lfloor t\rfloor$, one gets

$$\frac{\log t}{\log\lfloor t\rfloor}\cdot\frac{\log D_\alpha^{K^t}(\mu)}{\log t}=\frac{\log D_\alpha^{K^t}(\mu)}{\log\lfloor t\rfloor}\le\frac{H_\alpha(X_{\lfloor t\rfloor})}{\log\lfloor t\rfloor}+\frac{\sqrt{n}t}{\lfloor t\rfloor\log\lfloor t\rfloor}.$$

Letting $t\to\infty$, one gets

$$\overline{\dim}^\alpha(\mu)\le\overline{d}^\alpha(X).$$

To show the other direction, fix $k$ and $t>0$, define $A_{i_1,\cdots,i_n,k,k}=[\frac{i_1}{k}-\frac{1}{k},\frac{i_1+1}{k}+\frac{1}{k})\times\cdots\times[\frac{i_n}{k}-\frac{1}{k},\frac{i_n+1}{k}+\frac{1}{k})$, and $p_{i_1,\cdots,i_n,k,k}=\mu(A_{i_1,\cdots,i_n,k,k})$. Then

$$\int\left(\int e^{-td(x,y)}\ \mathrm{d}\mu\right)^{\alpha-1}\ \mathrm{d}\mu=\sum_j\int_{A_{j_1,\cdots,j_n,k}}(\int e^{-td(x,y)}\ \mathrm{d}\mu)^{\alpha-1}\ \mathrm{d}\mu.$$

$$\le\sum_j\int_{A_{j_1,\cdots,j_n,k}}\left(\int_{A_{j_1,\cdots,j_n,k,k}}e^{-td(x,y)}\ \mathrm{d}\mu+e^{-t/k}\right)^{\alpha-1}\ \mathrm{d}\mu$$

$$\le\sum_j p_{j_1,\cdots,j_n,k}(p_{j_1,\cdots,j_n,k,k}+e^{-t/k})^{\alpha-1}.$$

Note that,

$$\sum_j p_{j_1,\cdots,j_n,k}\left(p_{j_1,\cdots,j_n,k,k}+e^{-t/k}\right)^{\alpha-1}=\sum_j p_{j_1,\cdots,j_n,k}\left(\sum_{l_1=j_1-1}^{j_1+1}\cdots\sum_{l_n=j_n-1}^{j_n+1}p_{l_1,\cdots,l_n,k}+e^{-t/k}\right)^{\alpha-1}$$

$$\leq C(\alpha)\left(\sum_j p_{j_1,\cdots,j_n,k}\left(\sum_{l_1=j_1-1}^{j_1+1}\cdots\sum_{l_n=j_n-1}^{j_n+1} p_{l_1,\cdots,l_n,k}^{\alpha-1}\right)+e^{-(\alpha-1)t/k}\right),$$

where $C(\alpha)$ is an absolute constant depending merely on $\alpha$. The key observation is the following: since $\alpha > 1$, by Hölder's inequality, one has

$$\sum_j p_{j_1,\cdots,j_n,k}\; p_{l_1,\cdots,l_n,k}^{\alpha-1} \leq \left(\sum_j p_{j_1,\cdots,j_n,k}^{\alpha}\right)^{\frac{1}{\alpha}}\left(\sum_j p_{l_1,\cdots,l_n,k}^{\alpha}\right)^{\frac{\alpha-1}{\alpha}} = \sum_j p_{j_1,\cdots,j_n,k}^{\alpha}.$$

Therefore, we have

$$\sum_j p_{j_1,\cdots,j_n,k}\left(p_{j_1,\cdots,j_n,k,k}+e^{-t/k}\right)^{\alpha-1} \leq C(\alpha)\left(3^n\sum_j p_{j_1,\cdots,j_n,k}^{\alpha}+e^{-(\alpha-1)t/k}\right).$$

Whence,

$$\frac{1}{1-\alpha}\log\int(\int e^{-td(x,y)}d\mu)^{\alpha-1}d\mu \geq C'(\alpha)+\frac{1}{1-\alpha}\log\left(3^n\sum_j p_{j_1,\cdots,j_n,k}^{\alpha}+e^{-(\alpha-1)t/k}\right),$$

where $C'(\alpha)$ is an absolute constant depending merely on $\alpha$. For $\epsilon > 0$, setting $t = k^{1+\epsilon}$, we have

$$(1+\epsilon)\cdot\frac{\log D_{\alpha}^{K^{k^{1+\epsilon}}}(\mu)}{\log k^{1+\epsilon}} \geq \frac{C'(\alpha)}{\log k}+\frac{\frac{1}{1-\alpha}\log\left(3^n\sum_j p_{j_1,\cdots,j_n,k}^{\alpha}+e^{-(\alpha-1)k^{\epsilon}}\right)}{\log k}.$$

Letting $k\to\infty$, we obtain

$$\overline{\dim}^{\alpha}(\mu) \geq \frac{\overline{d}^{\alpha}(X)}{(1+\epsilon)}.$$

On the other hand, setting $k = \lfloor t^{1-\epsilon}\rfloor$, one has

$$\frac{\log D_{\alpha}^{K^t}(\mu)}{\log t} \geq \frac{C'(\alpha)}{\log t}+\frac{\frac{1}{1-\alpha}\log\left(3^n\sum_j p_{j_1,\cdots,j_n,\lfloor t^{1-\epsilon}\rfloor}^{\alpha}+e^{-(\alpha-1)t^{\epsilon}}\right)}{\log\lfloor t^{1-\epsilon}\rfloor}\cdot\frac{\log\lfloor t^{1-\epsilon}\rfloor}{\log t}.$$

Letting $t \to \infty$, we get,

$$\underline{\dim}^{\alpha}(\mu) \geq (1-\epsilon)\underline{d}^{\alpha}(X).$$

Since $\epsilon > 0$ is arbitrary, the desired result now follows.

For $\alpha \in (0,1)$, on one hand, we could repeat the argument above to show that

$$\frac{1}{1-\alpha} \log \int \left( \int e^{-td(x,y)} \, \mathrm{d}\mu \right)^{\alpha-1} \, \mathrm{d}\mu \leq H_{\alpha}(X_k) + \sqrt{n}t/k,$$

which implies that

$$\underline{\dim}^{\alpha}(\mu) \leq \underline{d}^{\alpha}(X), \quad \overline{\dim}^{\alpha}(\mu) \leq \overline{d}^{\alpha}(X).$$

On the other hand, given $L > 0$, for $\alpha \in (0,1)$, we have

$$\frac{1}{1-\alpha} \log \int \left( \int e^{-td(x,y)} d\mu \right)^{\alpha-1} d\mu \geq \frac{1}{1-\alpha} \log \left( \sum_{j=-kL}^{kL} p_{j_1,\cdots,j_n,k} \left( p_{j_1,\cdots,j_n,k,k} + e^{-t/k} \right)^{\alpha-1} \right).$$

By the reversed Hölder's inequality, we have

$$\sum_{j=-kL}^{kL} p_{j_1,\cdots,j_n,k} \left( p_{j_1,\cdots,j_n,k,k} + e^{-t/k} \right)^{\alpha-1}$$

$$\geq \left( \sum_{j=-kL}^{kL} p^{\alpha}_{j_1,\cdots,j_n,k} \right)^{\frac{1}{\alpha}} \left( \sum_{j=-kL}^{kL} \left( p_{j_1,\cdots,j_n,k,k} + e^{-t/k} \right)^{\alpha} \right)^{\frac{\alpha-1}{\alpha}}$$

Using the elementary inequality $(x+y)^{\alpha} \leq x^{\alpha} + y^{\alpha}$, for $\alpha \in (0,1)$ and $x, y \geq 0$, one has

$$\left( \sum_{j=-kL}^{kL} \left( p_{j_1,\cdots,j_n,k,k} + e^{-t/k} \right)^{\alpha} \right)^{\frac{\alpha-1}{\alpha}} \geq \left( \sum_{j=-kL}^{kL} \left( \sum_{l_1=j_1-1}^{j_1+1} \cdots \sum_{l_n=j_n-1}^{j_n+1} p^{\alpha}_{l_1,\cdots,l_n,k} + e^{-\alpha t/k} \right) \right)^{\frac{\alpha-1}{\alpha}}$$

$$\geq \left( 3^n \sum_{j} p^{\alpha}_{j_1,\cdots,j_n,k} + (2kL+1)^n e^{-\alpha t/k} \right)^{\frac{\alpha-1}{\alpha}}.$$

Therefore, we have

$$\sum_{j=-kL}^{kL} p_{j_1,\cdots,j_n,k}\left(p_{j_1,\cdots,j_n,k,k}+e^{-t/k}\right)^{\alpha-1}$$

$$\geq \left(\sum_{j=-kL}^{kL} p^{\alpha}_{j_1,\cdots,j_n,k}\right)^{\frac{1}{\alpha}}\left(3^n\sum_j p^{\alpha}_{j_1,\cdots,j_n,k}+(2kL+1)^n e^{-\alpha t/k}\right)^{\frac{\alpha-1}{\alpha}}$$

For $\epsilon>0$, choose $L$ large enough such that $\text{supp}(\mu)\subset[-L,L]^n$. Then

$$\log D^{K^t}_{\alpha}(\mu)\geq \frac{1}{\alpha(1-\alpha)}\log\sum_j p^{\alpha}_{j_1,\cdots,j_n,k}-\frac{1}{\alpha}\log\left(3^n\sum_j p^{\alpha}_{j_1,\cdots,j_n,k}+(2kL+1)^n e^{-\alpha t/k}\right).$$

By a similar argument as in the case $\alpha>1$, the conclusion follows. □

### 4.2 Maximum diversity of a compact metric space

In this section, we will specialize to the case of a compact metric space equipped with a metric-based similarity kernel. In continuation of existing literature, the underlying viewpoint will be to develop maximum diversity as a measure of size for compact metric spaces. Just as the maximum Rényi entropy of order $\alpha$ among all densities supported on a fixed compact subset of $\mathbb{R}^n$ is independent of $\alpha$, maximization of diversity is also independent of the order.

**Theorem 4.2.1** (Leinster-Meckes-Roff Maximization theorem)**.** *[64, 66] Let $X$ be a compact Hausdorff space and $K$ a symmetric similarity kernel on $K$. There exists a measure $\mu^*\in\mathcal{P}(X)$ which maximizes $D^K_{\alpha}(\cdot)$ for all $\alpha\in[0,\infty]$. Moreover, the "diversity profile" $\alpha\mapsto D^K_{\alpha}(\mu^*)$ for $\mu^*$ is constant. Consequently, the maximum diversity $\sup_{\mu\in\mathcal{P}(X)} D^K_{\alpha}(\mu)$ is independent of $\alpha$.*

Therefore, we can define the maximum diversity of a compact space with similarities by using any order $\alpha$.

**Definition 17** (Maximum diversity)**.** Let $(X, K)$ be a compact space with similarities. The maximum diversity of $(X, K)$ is defined by,

$$D_{\max}^K(X) = \sup_{\mu \in \mathcal{P}(X)} D_\alpha^K(\mu) \in (0, \infty),$$

for any $\alpha \in [0, \infty]$. In particular,

$$D_{\max}^K(X) = \sup_{\mu \in \mathcal{P}(X)} \frac{1}{\int \int K(x, y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)}.$$

**Remark 4.2.2.** *By construction, maximum diversity is monotone with respect to set containment.*

**Remark 4.2.3.** *When compact Hausdorff space $X$ is a metric space and the kernel $K$ is a function of the metric $d$, for example when $K$ is the Laplace or the Gaussian kernel, the maximum diversity $D_{\max}^K(X)$ becomes an isometry-invariant. While developing maximum diversity as a notion of "size" of a compact metric space, we will be primarily interested in the Laplace kernel on the compact metric space, in which case the maximum diversity will be denoted by $|X|_+$.*

We now explain why maximum diversity is denoted by $|\cdot|_+$ when calculated with respect to the Laplace kernel. Maximum diversity for the Laplace kernel is related to another invariant of compact metric spaces called magnitude (see [63]). Magnitude behaves well when restricted to compact positive definite metric spaces, i.e. metric spaces $(X, d)$ such that the corresponding Laplace kernel is positive definite:

$$\sum_{1 \leq i,j \leq n} e^{-d(x_i, x_j)} \eta_i \eta_j \geq 0$$

for all $x_1, \cdots, x_n \in X, \eta_1, \cdots, \eta_n \in \mathbb{R}$. Examples of positive definite metric spaces include $\mathbb{R}^n$ with the metric induced by $p$-norms, when $1 \leq p \leq 2$, all spheres $\mathbb{S}^l$ with the geodesic metric, real and complex hyperbolic spaces, etc. For compact positive definite metric spaces $(X, d)$, magnitude can be defined as (see [63, 80]).

$$|X| = \sup_{\mu \in \mathcal{M}(X)} \frac{\mu(X)^2}{\int\int e^{-d(x,y)} \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y)},$$

where $\mathcal{M}(X)$ is the space of all (finite) signed Borel measures on $X$. One notices that when the supremum is restricted to positive measures, maximum diversity is obtained. That is why we denote maximum diversity under the Laplace kernel by $|\cdot|_+$.

**From now on, unless stated otherwise, the maximum diversity of a compact metric space $(X, d)$ will always mean the maximum diversity as measured under the Laplace kernel and will be denoted by $|X|_+$.** Under the assumption of positive-definiteness, maximum diversity enjoys several additional properties such as continuity with respect to the Gromov-Hausdorff metric [79, Proposition 2.11] and uniqueness of the maximum diversity achieving measure when the space is finite [66, Proposition 8.7].

We now proceed to prove a subadditivity property of maximum diversity, via its entropic version.

**Definition 18.** Denote by $2^{[n]}$ the collection of all subsets of $[n] = \{1, \dots, n\}$. A function $\beta : 2^{[n]} \to \mathbb{R}_+$ is called a fractional partition if for each $i \in [n]$, we have $\sum_{s \in 2^{[n]}} \beta(s) \mathbf{1}_s(i) = 1$.

Recall that fractional subadditivity, that we prove below, reduces to the usual subadditivity when $\beta$ is chosen such that $\beta(\{i\}) = 1$ for all singletons $\{i\}$ and 0 otherwise.

**Proposition 4.2.4** (Fractional subadditivity of diversity)**.** *Let $(X, d)$ be a positive definite metric space, and $\mu_i$, $i \in [n]$ be probability measures on $X$ such that* $\mathrm{supp}(\mu_i) \cap \mathrm{supp}(\mu_j) = \emptyset$*, and $\beta$ be a fractional partition function. Let $\mu$ be the mixture of the $\mu_i$, i.e. $\mu = \sum_{i=1}^n \lambda_i \mu_i$, $\sum_{i=1}^n \lambda_i = 1$, $\lambda_i \geq 0$, for each $i$. Then for $\alpha \in [0, \infty]$,*

$$D_\alpha(\mu) \leq \sum_{s \in \mathcal{C}} \beta(s) D_\alpha(\mu_s),$$

*where* $\mu_s = \frac{\sum_{i\in s}\lambda_i\mu_i}{\sum_{i\in s}\lambda_i}$, *and* $\mathcal{C} = 2^{[n]}$.

*Proof.* For $\alpha \in [0,1)\cup(1,\infty)$, we have

$$D_\alpha(\mu) = \left(\int (K\mu)^{\alpha-1}\,\mathrm{d}\mu\right)^{1/(1-\alpha)} = \left(\sum_{i\in[n]}\lambda_i\int (K\mu)^{\alpha-1}\,\mathrm{d}\mu_i\right)^{1/(1-\alpha)}$$

$$= \left(\sum_{i\in[n]}\left(\sum_{s\in\mathcal{C}:i\in s}\beta(s)\right)\lambda_i\int (K\mu)^{\alpha-1}\,\mathrm{d}\mu_i\right)^{1/(1-\alpha)}$$

$$= \left(\sum_{s\in\mathcal{C}}\beta(s)\sum_{i\in s}\lambda_i\int (K\mu)^{\alpha-1}\,\mathrm{d}\mu_i\right)^{1/(1-\alpha)}$$

$$\leq \left(\sum_{s\in\mathcal{C}}\beta(s)\sum_{i\in s}\lambda_i\int (\sum_{i\in s}\lambda_i K\mu_i)^{\alpha-1}\,\mathrm{d}\mu_i\right)^{1/(1-\alpha)}$$

$$= \left(\sum_{s\in\mathcal{C}}\beta(s)(\sum_{i\in s}\lambda_i)^{\alpha}\int (K\mu_s)^{\alpha-1}\,\mathrm{d}\mu_s\right)^{1/(1-\alpha)}.$$

Note that for $\alpha \in [0,1)\cup(1,\infty)$, $x^{1/(1-\alpha)}$ is convex, and

$$\sum_{s\in\mathcal{C}}\beta(s)\sum_{i\in s}\lambda_i = 1.$$

Therefore,

$$\left(\sum_{s\in\mathcal{C}}\beta(s)(\sum_{i\in s}\lambda_i)^{\alpha}\int (K\mu_s)^{\alpha-1}\,\mathrm{d}\mu_s\right)^{1/(1-\alpha)}$$

$$= \left(\sum_{s\in\mathcal{C}}\beta(s)(\sum_{i\in s}\lambda_i)(\sum_{i\in s}\lambda_i)^{\alpha-1}\int (K\mu_s)^{\alpha-1}\,\mathrm{d}\mu_s\right)^{1/(1-\alpha)}$$

$$\leq \sum_{s\in\mathcal{C}}\beta(s)\left(\int (K\mu_s)^{\alpha-1}\,\mathrm{d}\mu_s\right)^{1/(1-\alpha)} = \sum_{s\in\mathcal{C}}\beta(s)D_\alpha(\mu_s).$$

For $\alpha = 1$ and $\infty$, take limits to get desired results.

□

In general, maximum diversity does not satisfy the principle of inclusion-exclusion on compact subsets of a given metric space. However, it is fractionally subadditive.

**Corollary 4.2.5** (Fractional subadditivity of maximum diversity)**.** *Let $(X, d)$ be a positive definite metric space, $A_1, A_2, \cdots, A_n \subseteq X$ be compact subsets, and $\beta : 2^{[n]} \to [0, \infty)$ be a fractional partition function. Then*

$$\left|\bigcup_{i=1}^{n} A_i\right|_+ \leq \sum_{s \in 2^{[n]}} \beta(s) \left|\bigcup_{i \in s} A_i\right|_+ .$$

*Proof.* Note that the corollary is straightforward from the previous proposition when the $A_i$ are disjoint. Now suppose $F \subseteq \cup_{i\in[n]} F_i$ and all sets are finite. Let $\tilde{F}_i = F_i \setminus \cup_{k=1}^{i-1} F_i$ denote the "disjointification" of the cover. Then $F \subseteq \cup_{i\in[n]} \tilde{F}_i$, as well as, $\cup_{i\in s} \tilde{F}_i \subseteq \cup_{i\in s} F_i$ for every $s \in 2^{[n]}$. By the corollary for disjoint sets and monotonicity of maximum diversity with respect to inclusion,

$$|F|_+ \leq \sum_{s\in 2^{[n]}} \beta(s) | \cup_{i\in s} \tilde{F}_i|_+ \leq \sum_{s\in 2^{[n]}} \beta(s) | \cup_{i\in s} F_i|_+.$$

Finally, consider compact sets $A, A_1, \ldots, A_n$ such that $A \subseteq \cup_i A_i$. For any fixed finite $F \subseteq A$, set $F_i = A_i \cap F$. Then $F \subseteq F_i$ and so,

$$|F|_+ \leq \sum_{s\in 2^n} \beta(s) | \cup_{i\in s} F_i|_+ \leq \sum_{s\in 2^n} \beta(s) | \cup_{i\in s} A_i|_+.$$

Taking supremum over all finite $F \subseteq A$ finishes the proof. □

We "scaled" a contoured kernel earlier to recover Rényi entropic notions. Similarly, we will scale the metric (which exactly corresponds to scaling of the Laplace kernel) to recover classical geometric quantities. In the following, $|t \cdot X|_+$ will denote the maximum diversity of the metric space $(X, td)$, for every $t > 0$. Before proceeding with our investigations, we will motivate and state some previous work that we use.

Let us consider the following heuristic. At $t = \infty$ scaling of the metric we obtain exponentiated discrete Rényi entropies from diversities in the finite case and exponentiated continuous Rényi entropy from (appropriately scaled) diversities in the continuous case. The maximum diversity in these cases therefore must yield maximum exponentiated Rényi entropies, which is the cardinality for the former and volume for the latter. This indeed does happen, we note it as the following proposition.

**Proposition 4.2.6** (Recovery of cardinality and volume)**.** *[63, 66, 80] Let $X$ be a compact subspace of $\ell_2^n$. If $X$ is a finite subspace, then*

$$\lim_{t\to\infty} |t \cdot X|_+ = \#(X).$$

*If $X$ is a subspace with nonzero volume $\lambda(X)$, then*

$$\lim_{t\to\infty} \frac{|t \cdot X|_+}{t^n} = \frac{\lambda(X)}{n!\omega_n},$$

*where $\omega_n$ is the volume of the unit ball in $\mathbb{R}^n$. In short, maximum diversity detects cardinality and volume.*

From the limiting expression above which recovers volume, it follows that maximum diversity detects ambient dimension for compact sets $X$ of positive measure via

$$\lim_{t\to\infty} \frac{\log |t \cdot X|_+}{\log t} = n.$$

Meckes [80] showed that the above holds in wider generality. To state his result, we first recall the definition of dimension appropriate for the context.

**Definition 19.** The Minkowski dimension of a compact metric space $(X, d)$ is defined by,

$$\dim_{\text{Mink}}(X) = \lim_{\epsilon\to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)},$$

provided that the limit exists. Here the covering number $N(\epsilon)$ is the minimal number of open balls of radius $\epsilon$ that are required to cover the space.

**Proposition 4.2.7** (Recovery of Minkowski dimension)**.** *[80, Theorem 7.1] Let $X$ be a compact metric space, and let $\dim_{Mink}(X)$ denote the Minkowski dimension of $X$. Then*

$$\lim_{t\to\infty} \frac{\log |t \cdot X|_+}{\log t} = \dim_{\text{Mink}}(X),$$

*and the left-hand side is defined if and only if the right-hand side is defined. In other words, the Minkowski dimension of $X$ is the asymptotic growth rate of $|t \cdot X|_+$.*

As our next step, we prove a generalization of the 1-dimensional Brunn-Minkowski inequality.

**Theorem 4.2.8** (Brunn-Minkowski for maximum diversity in $\mathbb{R}$)**.** *Let $A, B$ be two compact sets in $\mathbb{R}$, and $\lambda \in (0, 1)$. Then*

$$|(1-\lambda)A + \lambda B|_+ \geq (1-\lambda)|A|_+ + \lambda|B|_+.$$

*Proof.* The proof is a consequence of two simple observations:

1. Maximum diversity satisfies the Cauchy-Davenport inequality for compact subsets $A, B \subseteq \mathbb{R}$:
$$|A + B|_+ \geq |A|_+ + |B|_+ - 1.$$
2. The function $\lambda \mapsto |\lambda \cdot A|_+$ is concave for every finite subset $A \subseteq \mathbb{R}$.

Both these observations directly follow from [65, Theorem 4.1] [1]. The second observation in turn implies the concavity of the function $\lambda \mapsto |\lambda \cdot B|_+ - (\lambda|B|_+ + (1-\lambda))$. This leads to inequalities

$$|(1-\lambda) \cdot A|_+ \geq (1-\lambda)|A|_+ + \lambda, |\lambda \cdot B|_+ \geq \lambda|B|_+ + (1-\lambda),$$

since a concave function of $\lambda$ which equals zero at the endpoints $\lambda = 0, \lambda = 1$ must stay non-negative in the interval $[0, 1]$.

[1] The theorem, as stated, is for another invariant of compact metric spaces called Magnitude. However, magnitude and maximum diversity are equivalent for compact subsets of $\mathbb{R}$ (see [65, 80]).

For the final step, we apply the Cauchy-Davenport inequality for maximum diversity to the sets $(1-\lambda)A$ and $\lambda B$,

$$
\begin{aligned}
|(1-\lambda)A + \lambda B|_+ &\geq |(1-\lambda)A|_+ + |\lambda B|_+| - 1 \\
&\geq (1-\lambda)|A|_+ + \lambda + \lambda|B|_+ + (1-\lambda) - 1 \\
&= (1-\lambda)|A|_+ + \lambda|B|_+.
\end{aligned}
$$

□

**Corollary 4.2.9** (Brunn-Minkowski inequality in $\mathbb{R}$)**.** *Let $A, B$ be compact subsets of $\mathbb{R}$ and $\lambda \in [0,1]$. Then,*

$$\mathrm{Vol}((1-\lambda)A + \lambda B) \geq (1-\lambda)\mathrm{Vol}(A) + \lambda \mathrm{Vol}(B).$$

A similar technique does not appear to work in higher dimensions under any norm that we considered. We have been unable to prove an exact analogue of the Brunn-Minkowski inequality in higher dimensions, even in its dimension-free form:

$$\mathrm{Vol}((1-\lambda)A + \lambda B) \geq \mathrm{Vol}(A)^{1-\lambda}\mathrm{Vol}(B)^{\lambda},$$

for compact sets $A, B \subseteq \mathbb{R}^n$. However one can show that the maximum diversity Brunn-Minkowski inequality is weakest under the Euclidean norm. We first prove a preliminary result towards this.

**Proposition 4.2.10** (Behaviour under contractions)**.** *Let $(X, d_1)$ and $(Y, d_2)$ be two compact metric spaces, and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Assume $T$ is a $\beta$-Lipschitz map from $X$ to $Y$ that push-forward $\mu$ to $\nu$. Then,*

$$D_\alpha^{K_1^\beta}(\mu) \geq D_\alpha^{K_2}(\nu).$$

*In particular, if $\mu \in \mathcal{P}(X)$ and $\nu$ is a maximum diversity achieving measure for $Y$,* *then*

$$|\beta \cdot X|_+ \geq |Y|_+.$$

*Proof.* Let us first assume $\alpha \in (-\infty, 1) \cup (1, \infty)$,

$$D_\alpha^{K_1^\beta}(\mu) = \left( \int_X \left( \int_X e^{-\beta d_1(x_1,x_2)} d\mu(x_2) \right)^{\alpha-1} d\mu(x_1) \right)^{1/(1-\alpha)}$$

$$\geq \left( \int_X \left( \int_X e^{-d_2(T(x_1),T(x_2))} d\mu(x_2) \right)^{\alpha-1} d\mu(x_1) \right)^{1/(1-\alpha)} \quad \text{(since } T \text{ is } L\text{-Lipschitz)}$$

$$= \left( \int_Y \left( \int_Y e^{-d_2(y_1,y_2)} d\nu(y_2) \right)^{\alpha-1} d\nu(y_1) \right)^{1/(1-\alpha)} = D_\alpha^{K_2}(\nu).$$

For $\alpha = \pm\infty, 1$, taking limits to complete the proof. □

**Remark 4.2.11.** *By the above proposition, whenever a probability measure $\nu$ on a metric space $(Y, d')$ is a pushforward of a probablity measure $\mu$ on $(X, d)$ under a 1-Lipschitz map, one can immediately conclude that diversities of all orders of $\nu$ are smaller than those of $\mu$. For example, every log-concave density with a 1-uniformly convex potential is the pushforward of the Gaussian under a 1-Lipschitz map [27] and so all its diversities are smaller than those of the Gaussian of the same order.*

As an easy application of Proposition 4.2.10, we show that in $\mathbb{R}$, among all compact sets of fixed volume, the line segments have the smallest maximum diversity.

**Corollary 4.2.12** (Isoperimetric inequality in $\mathbb{R}$)**.** *Let $A$ be a compact set in $\mathbb{R}$, then*

$$|A|_+ \geq |A^*|_+,$$

*where $A^*$ is the closed interval centered at the origin having the same volume as $A$. The inequality holds if and only if $A$ is a translation of $A^*$.*

*Proof.* Assume that $A^* = [-a, a]$. Using the continuity of maximum diversity with respect to the Gromov-Hausdorff metric, one may choose an partition $P_n = \{-a = x_0, x_1, \ldots, x_n = a\}$ of $A^*$ such that for any given $\varepsilon > 0$, one has that

$$|P_n|_+ \geq |A^*|_+ - \varepsilon.$$

Assume that $\min A = b = x'_0,\ \max A = c = x'_n$. Then one may find points $x'_i$ in $A$ such that

$$x_{i+1} - x_i = \mathrm{Vol}_1((x'_{i+1} - x'_i) \cap A).$$

Define a map $T : P'_n = \{x'_0, x'_1, \ldots, x'_n\} \mapsto \mathbb{R}$ by mapping $x'_i$ to $x_i$. It is clear that $T$ is a 1-Lipschitz map. Therefore, by Proposition 4.2.10 and monotonicity of maximum diversity, one has

$$|A|_+ \geq |P'_n|_+ \geq |P_n|_+ \geq |A^*|_+ - \varepsilon.$$

The desired result now follows. □

As an easy consequence of the change of variables, we also have the following.

**Lemma 4.2.13.** *Let $\|\cdot\|_Q$ be an Euclidean norm on $\mathbb{R}^n$ induced by a positive quadratic form $Q$, and $A$ be a compact subset of $\mathbb{R}^n$. Define $K_Q(x, y) := e^{-\sqrt{\langle Q(x-y),(x-y)\rangle}}$, then*

$$D^{K_Q}_{\max}(A) = D^{K}_{\max}(T(A)),$$

*where $K$ is the kernel induced by the standard Euclidean norm and $T = Q^{1/2}$.*

*Proof.* Since $T \in GL(\mathbb{R}^n)$, one has

$$\sup_{\mu \in \mathcal{P}(A)} \frac{1}{\int\int e^{-\|Tx-Ty\|_2}\, \mathrm{d}\mu(x)\, \mathrm{d}\mu(y)} = \sup_{\nu \in \mathcal{P}(T(A))} \frac{1}{\int\int e^{-\|x-y\|_2}\, \mathrm{d}\nu(x)\, \mathrm{d}\nu(y)}.$$

□

As discussed earlier, we will show that the Euclidean maximum diversity Brunn-Minkowski inequality is the weakest among the corresponding results for all norms. In fact, one can show such a result for a wide class of geometric inequalities. However, for simplicity we restrict ourselves to inequalities which bound below the size of the sumset in terms of the summands.

**Theorem 4.2.14.** *Suppose* $(X_n, \|\cdot\|_{X_n})_{n\geq 1}$ *is a sequence of Banach spaces such that* $\dim X_n \to \infty$*, as* $n \to \infty$*. Let* $\Psi(\cdot)$ *be an non-decreasing continuous function on* $[0,\infty)$ *and* $\Phi(\cdot,\cdot)$ *be coordinate-wise non-decreasing continuous on* $[0,\infty)\times[0,\infty)$*. Assume that the inequality is true in each* $X_n$*: For all compact (or compact and convex) sets* $A, B$*,*

$$\Psi(|sA+tB|_{+,X_n}) \geq \Phi(|A|_{+,X_n}, |B|_{+,X_n}).$$

*Then the inequality: For each* $n$*, for all compact (or compact and convex) sets* $A, B \subset \ell_2^n$*,*

$$\Psi(|sA+tB|_{+,2}) \geq \Phi(|A|_{+,2}, |B|_{+,2}),$$

*is true.*

*Proof.* The key ingredient of the proof is Dvoretzy's theorem, which, given $\epsilon > 0, k \in \mathbb{N}$, guarantees the existence of a number $N(k,\epsilon)$ and a positive quadratic form $Q$ on $\mathbb{R}^k$ such that every Banach space $X$ of dimension $\geq N(\epsilon, k)$ contains a subspace $E$ of dimension $k$ such that the norm $\|\cdot\|_Q$ induced by by the quadratic form $Q$ on $E$ satisfies

$$\|x\|_Q \leq \|x\|_X \leq (1+\epsilon)\|x\|_Q.$$

For the next step, we use Proposition 4.2.10 to conclude that

$$|A|_{+,Q} \leq |A|_{+,X} \leq |(1+\epsilon)\cdot A|_{+,Q}$$

for every compact $A \subseteq E$. Suppose that the sumset maximum diversity inequality holds in $X$, and $A, B$ are compact subsets of $E$. Then

$$\Psi(|(1+\epsilon)\cdot(sA+tB)|_{+,Q}) \geq \Psi(|sA+tB|_{+,X}) \geq \Phi(|A|_{+,X}, |B|_{+,X}) \geq \Phi(|A|_{+,Q}, |B|_{+,Q}).$$

By 4.2.13, for all compact (or compact convex) $A, B$,

$$\Psi(|(1+\epsilon)\cdot(sA+tB)|_{+,2}) \geq \Phi(|A|_{+,2}, |B|_{+,2}).$$

Finally, by using the sequence of Banach spaces $X_n$ with dimension growing to $\infty$ we can send $\epsilon \to 0$, and invoking the continuity of maximum diversity we obtain the result for all $\ell_2^n$. □

## 4.3 Behaviour of diversity along geodesics in the Wasserstein space and consequences

Our first result shows that on $\mathcal{P}(\mathbb{R})$, the entropy $H_\alpha(\mu)$ defined in section 2 is displacement concave for $\alpha \geq 1$. In the sequel, if a random variable $X$ has law $\mu$, we also use $H_\alpha(X)$ to represent the entropy.

**Proposition 4.3.1.** *Let $\mu, \nu$ be two probability measures on $\mathbb{R}$, and $U$ be the uniform distribution on $[0,1]$. Let $f$ and $g$ be two monotone functions pushing forward the Lebesgue measure on $[0,1]$ to the probability measures $\mu$ and $\nu$, respectively. Then for $\lambda \in (0,1)$ and $\alpha \geq 1$, we have*

$$H_\alpha((1-\lambda)f(U)+\lambda g(U)) \geq (1-\lambda)H_\alpha(\mu)+\lambda H_\alpha(\nu).$$

*Proof.* Let $U'$ be an independent copy of $U$.
For $\alpha > 1$, by definition, we have

$$H_\alpha((1-\lambda)f(U)+\lambda g(U)) = \frac{1}{1-\alpha}\log\left(\mathbb{E}\left(\mathbb{E}e^{-|(1-\lambda)(f(U)-f(U'))+\lambda(g(U)-g(U'))|}\right)^{\alpha-1}\right).$$

Since $f(x)$ and $g(x)$ are increasing, we have

$$|(1-\lambda)(f(U)-f(U'))+\lambda(g(U)-g(U'))| = (1-\lambda)|f(U)-f(U')|+\lambda|g(U)-g(U')|.$$

By using Hölder's inequality twice, we obtain

$$H_\alpha((1-\lambda)f(U)+\lambda g(U))$$

$$\geq \frac{1-\lambda}{1-\alpha}\log\mathbb{E}\left(\mathbb{E}e^{-|f(U)-f(U')|}\right)^{\alpha-1}+\frac{\lambda}{1-\alpha}\log\mathbb{E}\left(\mathbb{E}e^{-|g(U)-g(U')|}\right)^{\alpha-1}$$

$$=(1-\lambda)H_\alpha(\mu)+\lambda H_\alpha(\nu).$$

For $\alpha = 1$, by definition, we have

$$H_1((1-\lambda)f(U)+\lambda g(U)) = -\mathbb{E}\left(\log\left(\mathbb{E}e^{-|(1-\lambda)(f(U)-f(U'))+\lambda(g(U)-g(U'))|}\right)\right).$$

The conclusion follows from Hölder's inequality.

□

Proposition 4.1 can be used to recover a BM inequality for maximum diversity, which is weaker than Theorem 3.11.

**Corollary 4.3.2.** *Let $A$ and $B$ be two compact sets in $\mathbb{R}$ and $0 < \lambda < 1$, then*

$$|(1-\lambda)A+\lambda B|_+ \geq |A|_+^{1-\lambda}|B|_+^{\lambda}.$$

*Proof.* Let $\mu$ and $\nu$ be the maximum diversity achieving probability measures for $A$ and $B$, respectively. By Theorem 4.2.1, we have

$$|A|_+ = D_2(\mu) \qquad |B|_+ = D_2(\nu).$$

By Proposition 4.3.1, one has

$$|(1-\lambda)A+\lambda B|_+ \geq D_2((1-\lambda)f(U)+\lambda g(U)) \geq D_2(\mu)^{1-\lambda}D_2(\nu)^{\lambda},$$

as desired. □

We emphasize that Proposition 4.3.1 does not hold on $\mathbb{R}^n$ with $n \geq 2$, therefore the BM inequality for maximum diversity in higher dimensional space does not follow from it. However, we still have the following:

**Proposition 4.3.3** (Laplace kernel)**.** *Let $A$ and $B$ be two compact sets in $\mathbb{R}^n$ equipped with Euclidean norm. Then for $\lambda \in (0,1)$, we have*

$$|(1-\lambda)A + \lambda B|_+ \geq |cA|_+^{1-\lambda} |cB|_+^{\lambda},$$

*where $c = \frac{\sqrt{2}}{2}$.*

*Proof.* Let $\mu$ and $\nu$ be the maximum diversity achieving probability measures for $A$ and $B$, respectively, such that

$$|A|_+ = D_2(\mu), \qquad |B|_+ = D_2(\nu).$$

Let $\mu_t = \gamma_n^t * \mu$, where $\gamma_n^t$ is the Gaussian density in $\mathbb{R}^n$ with covariance matrix $tI_n$; it is routine to check that $\mu_t$ converges to $\mu$ weakly as $t \searrow 0$. Now, let $\nabla\psi_t$ be the optimal transport map from $\mu_t$ to $\nu$, Note that for $x, y \in \mathbb{R}^n$, by monotonicity of $\nabla\psi_t$, we have

$$\langle x - y, \nabla\psi_t(x) - \nabla\psi_t(y)\rangle \geq 0.$$

Therefore

$$\|(x-y) + (\nabla\psi_t(x) - \nabla\psi_t(y))\|_2 \geq \frac{\|x-y\|_2 + \|\nabla\psi_t(x) - \nabla\psi_t(y)\|_2}{\sqrt{2}}.$$

Now, repeating the proof in Proposition 4.1, one obtains

$$D_2\left([(1-\lambda)x + \lambda\nabla\psi_t(x)]\#\mu_t\right) \geq D_2(\mu_t)^{1-\lambda} D_2(\nu)^{\lambda}.$$

Recall that the optimal transference plan $\pi_t$ in the Kantorovich transportation problem between $\mu_t$ and $\nu$ for the quadratic cost function $c(x,y) = \|x-y\|^2$ is given by

$$\pi_t = (\mathrm{Id} \times \psi_t)\#\mu_t,$$

where Id is the identity map. Therefore, we have

$$D_2\left([(1-\lambda)x + \lambda y]\#\pi_t\right) \geq D_2(\mu_t)^{1-\lambda} D_2(\nu)^{\lambda}.$$

Since $A, B$ are compact, $\pi_t$ is a probability density with a bounded second moment for all $t \in (0,1]$, say. Therefore $(\pi_t)$ is tight, hence has a weak limit point $\pi$ as $t \searrow 0$. Since the marginals of $\pi_t$ converge to $\mu$ and $\nu$ separately, one deduces that $\pi$ is supported on $A \times B$.

Recall that diversity is continuous under weak topology, therefore, let $t \searrow 0$, one has

$$D_2\left([(1-\lambda)x + \lambda y]\#\pi\right) \geq D_2(\mu)^{1-\lambda} D_2(\nu)^{\lambda}.$$

The desired result now follows.

□

**Proposition 4.3.4** (Gaussian kernel)**.** *Let $X$ and $Y$ be two continuous random vectors taking values on $\mathbb{R}^n$ equipped with the Gaussian similarity kernel. Then for $\lambda \in (0,1)$, one has*

$$H_\alpha(\sqrt{1-\lambda}X + \sqrt{\lambda}Y) \geq (1-\lambda)H_\alpha(X) + \lambda H_\alpha(Y),$$

*where $Y = \nabla\psi(X)$, and $\nabla\psi$ is the optimal transport map from the law of $X$ to the law of $Y$.*

*Proof.* By definition, for $\alpha \in (-\infty, 1) \cup (1, \infty)$,

$$D_\alpha(\sqrt{1-\lambda}X + \sqrt{\lambda}Y) = \left(\mathbb{E}\left(\mathbb{E}e^{-\|\sqrt{1-\lambda}(X-X') + \sqrt{\lambda}(\nabla\psi(X) - \nabla\psi(X'))\|_2^2}\right)^{\alpha-1}\right)^{1/(1-\alpha)}.$$

By the monotonicity of $\nabla\psi$, we have, for $x, y \in \mathbb{R}^n$,

$$\langle x - y, \nabla\psi(x) - \nabla\psi(y)\rangle \geq 0.$$

Therefore,

$$\|\sqrt{1-\lambda}(X - X') + \sqrt{\lambda}(\nabla\psi(X) - \nabla\psi(X'))\|_2^2 \geq (1-\lambda)\|X - X'\|_2^2 + \lambda\|Y - Y'\|_2^2.$$

By Hölder's inequality, one has

$$\mathbb{E}e^{-\|\sqrt{1-\lambda}(X-X')+\sqrt{\lambda}(\nabla\psi(X)-\nabla\psi(X'))\|_2^2} \leq (\mathbb{E}e^{-\|X-X'\|_2^2})^{(1-\lambda)}(\mathbb{E}e^{-\|\nabla\psi(X)-\nabla\psi(X')\|_2^2})^{\lambda}.$$

Therefore,

$$D_\alpha(\sqrt{1-\lambda}X + \sqrt{\lambda}Y) \geq D_\alpha(X)^{(1-\lambda)}D_\alpha(Y)^{\lambda}.$$

For $\alpha = 1$, one has

$$H_1(\sqrt{1-\lambda}X + \sqrt{\lambda}Y) = -\left(\mathbb{E}\left(\log \mathbb{E}e^{-\|\sqrt{1-\lambda}(X-X')+\sqrt{\lambda}(\nabla\psi(X)-\nabla\psi(X'))\|_2^2}\right)\right).$$

Again by monotonicity of $\nabla\psi$ and Hölder's inequality, we have

$$H_1(\sqrt{1-\lambda}X + \sqrt{\lambda}Y) \geq (1-\lambda)H_1(X) + \lambda H_1(Y).$$

The proof is complete.

□

### 4.4 Relationship with the other notion of diversity

The defining metric $d : X \times X \to [0, \infty)$ of metric space $(X, d)$ can be equivalently described as a function on the set of two-element subsets of $X$ satisfying certain properties. In [25], with the aim of extending the notion of hyperconvexity, Bryant and Tupper generalize this viewpoint on metrics to admit functions taking any finite subset of $X$ as argument.

**Definition 20.** [25] Let $X$ be a a set, $\delta : \{F \in 2^X : \#F < \infty\} \to [0, \infty)$. The pair $(X, \delta)$ is called a diversity if $\delta$ satisfies the conditions:

1. $\delta(A) = 0$ if and only if $\#A \leq 1$;
2. $\delta(A \cup B) \leq \delta(A \cup C) + \delta(C \cup B)$, whenever $B \neq \emptyset$.

**Remark 4.4.1.** *We call this notion BT diversity to distinguish it from the notion of diversity for probability measures that is at the heart of our considerations in the present work. Interestingly, the authors of [25] termed their quantity diversity because a special case of their definition appears in the literature on phylogenetics and ecological diversity.*

For examples of BT diversities, we refer the reader to [24, 25]. In this section, we seek to extend this list of examples by showing that the maximum diversity of finite subsets of certain metric spaces naturally give rise to BT diversities.

Let $(X, d)$ be a metric space. Define a function

$$\delta(A) = \max\{|A|_+ - 1, 0\},$$

on all finite subsets $A \subseteq X$. Clearly, $\delta(A) = 0$ if and only if $\#A \leq 1$. Thus, the first requirement for $(X, \delta)$ to be a BT diversity is always satisfied. For the second requirement, we need maximum diversity to satisfy a property stronger than subadditivity (Corollary 4.2.5) but weaker than submodularity.

**Lemma 4.4.2.** *Let $(X, d)$ be a metric space. The pair $(X, \delta)$ is a BT diversity if and only if,*

$$|A \cup C|_+ + 1 \leq |A|_+ + |C|_+,$$

*for non-empty compact subsets $A, C \subseteq X$ that intersect in exactly one point.*

*Proof.* For the forward implication, simply pick $B = A \cap C$ when $A$ and $C$ intersect in exactly one point.

Now we proceed towards proving the backwards implication. Thus, we assume that $|A \cup C|_+ + 1 \le |A|_+ + |C|_+$ for non-empty compact subsets $A, C \subseteq X$ that intersect in exactly one point. For $(X, d)$ to be a diversity, the inequality $\delta(A \cup C) \le \delta(A \cup B) + \delta(B \cup C)$, for finite sets $A, B, C$ such that $B \neq \emptyset$, must be verified. This requirement is easily seen to be satisfied even if one of $A$ or $C$ is empty. So we may assume that both $A$ and $C$ are non-empty. Now, the inequality to be verified reads

$$|A \cup C|_+ + 1 \le |A \cup B|_+ + |B \cup C|_+$$

in the language of maximum diversity, since $\delta(\cdot) = |\cdot|_+ - 1$ for all the sets involved. First assume $A \cap C = \emptyset$. Fix $B \neq \emptyset$, and $x_0 \in B$. Then $A' = A \cup \{x_0\}$ and $C' = C \cup \{x_0\}$ intersect in exactly one point. Then

$$|A \cup C|_+ + 1 \le |A' \cup C'|_+ + 1 \le |A'|_+ + |C'|_+ \le |A \cup B|_+ + |B \cup C|_+$$

as desired. If $A \cap C \supseteq \{x_0\}$, then let $A' = A$, $C' = (C \setminus A) \cup \{x_0\}$ so that $A'$ and $C'$ again intersect exactly in one point. Now by the assumption for maximum diversity of sets intersecting in one point and monotonicity of maximum diversity under set-inclusion, we have $|A \cup C|_+ + 1 = |A' \cup C'|_+ + 1 \le |A'|_+ + |C'|_+ \le |A|_+ + |C|_+$. □

Recall that a pointed metric space $(X, d, x_0)$ is simply a metric space $(X, d)$ with a distinguished point $x_0$. The wedge sum of two pointed metric spaces is defined by gluing them at their distinguished points.

**Definition 21.** For two pointed metric spaces, $X_1 = (X_1, d_1, x_1), X_2 = (X_2, d_2, x_2)$ we define their wedge sum $X_1 \vee X_2$ as pointed metric space $(X, d, x_0)$ on the underlying set $X = X_1 \sqcup X_2 / x_1 \sim x_2$, equipped with the metric

$$d(x, y) = \begin{cases} d_i(x, y) & \text{if } x, y \in X_i, \\ d_i(x, x_i) + d_j(x_j, y) & \text{if } x \in X_i, y \in X_j, i \neq j, \end{cases}$$

with the distinguished point $x_0$ which is the equivalence class of the $x_i$.

Maximum diversity of pointed metric spaces (with respect to the Laplace kernel) can be calculated in the exact same way as for metric spaces, and the distinguished point plays no role. The following proposition tells us that for any metric space $(X, d)$, $(X, \delta)$ is indeed a BT diversity,

**Proposition 4.4.3.** *Suppose $(X, d)$ is a metric space such that all non-empty finite subsets are viewed as pointed metric spaces. Then, the corresponding $(X, \delta)$ is a BT diversity.*

*Proof.* By Lemma 4.4.2, we need to show that for finite subsets $A, B \subseteq X$ which intersect in exactly one point, say $z$, one has

$$|A \cup B|_+ + 1 \leq |A|_+ + |B|_+$$

The set-theoretic identity map $A \vee C \to A \cup C$ is then a contraction, hence by Proposition 4.2.10, one has

$$|A \vee B|_+ \geq |A \cup B|_+.$$

Therefore, it suffices to show that

$$|A \vee B|_+ + 1 \leq |A|_+ + |B|_+$$

To this end, let $\rho$ be a diversity-maximizing probability measure on $A \vee B$, and denote by $z$ the common point of $A$ and $B$ in $A \vee B$. We will consider two cases, according to whether $z \in \operatorname{supp} \rho$.

Suppose first that $z \in \operatorname{supp} \rho$. Then we can write $\rho = \mu + \nu + \varepsilon\delta_z$, where $\mu$ is supported on $A \setminus \{z\}$, $\nu$ is supported on $B \setminus \{z\}$, and $\varepsilon > 0$. We denote

$$a = \|\mu\|, \quad b = \|\nu\|, \quad M = Z\mu(z) = \int e^{-d(x,z)}\, d\mu(x), \quad N = Z\nu(z) = \int e^{-d(x,z)}\, d\nu(x),$$

and let $D = |A \vee B|_+$. We have $Z\rho(x) = 1/D$ for each $x \in \operatorname{supp}\rho$, which implies that

$$\frac{1}{D} = \begin{cases} Z\mu(x) + e^{-d(x,z)}(N+\varepsilon) = Z(\mu + (N+\varepsilon)\delta_z)(x) & \text{if } x \in \operatorname{supp}\mu, \\ Z\nu(x) + e^{-d(x,z)}(M+\varepsilon) = Z(\nu + (M+\varepsilon)\delta_z)(x) & \text{if } x \in \operatorname{supp}\nu, \\ M + N + \varepsilon & \text{if } x = z. \end{cases}$$

Now define $\widetilde{\mu} = \mu + (N+\varepsilon)\delta_z$ and $\widetilde{\nu} = \nu + (M+\varepsilon)\delta_z$. Then $Z\widetilde{\mu}(x) = M+N+\varepsilon = \frac{1}{D}$ for all $x \in \operatorname{supp}\widetilde{\mu} = (\operatorname{supp}\mu) \cup \{z\}$, and also $Z\widetilde{\nu}(x) = N + M + \varepsilon = \frac{1}{D}$ for all $x \in \operatorname{supp}\widetilde{\nu} = (\operatorname{supp}\nu) \cup \{z\}$. Thus $\widetilde{\mu}$ and $\widetilde{\nu}$ are scalar multiples of weight measures for their supports, and the diversity-maximizing measures for those supports are their mass-one normalizations. Therefore

$$|A|_+ + |B|_+ \geq D\,\|\widetilde{\mu}\| + D\,\|\widetilde{\nu}\| = D(a + N + \varepsilon + b + M + \varepsilon) = D + 1$$

since $a + b + \varepsilon = 1$ and $M + N + \varepsilon = 1/D$.

Now suppose that $z \notin \operatorname{supp}\rho$. In this case we can similarly write $\rho = \mu + \nu$, and we have

$$\frac{1}{D} = \begin{cases} Z\mu(x) + e^{-d(x,z)}N = Z(\mu + N\delta_z)(x) & \text{if } x \in \operatorname{supp}\mu, \\ Z\nu(x) + e^{-d(x,z)}M = Z(\nu + M\delta_z)(x) & \text{if } x \in \operatorname{supp}\nu. \end{cases}$$

Define $\widetilde{\mu} = \mu + N\delta_z$ and $\widetilde{\nu} = \nu + M\delta_z$. We now estimate

$$D_1(\widetilde{\mu}/(a+N)) = \int (Z\widetilde{\mu})^{-1}\,d\widetilde{\mu} = Da + N(Z\widetilde{\mu})(z)^{-1} = Da + \frac{N}{M+N},$$

and similarly

$$D_1(\widetilde{\nu}/(b+M)) \geq Db + \frac{M}{M+N}.$$

Therefore

$$|A|_+ + |B|_+ \geq D_1(\widetilde{\mu}/(a+N)) + D_1(\widetilde{\nu}/(b+M)) \geq D + 1$$

since $a + b = 1$.

$\square$

We remark that submodularity more generally would be interesting to prove, as it has found remarkable utility not only in its classical domain of combinatorics [47], but also in information theory (see, e.g., [46, 76, 106]), additive combinatorics (see, e.g., [73, 74, 87]), and convex geometry (see, e.g., [42, 44]).

# Chapter 5

# CONCLUSION

In the previous chapters, we explore the connections between information theory and other topics. Apart from those results, there are many interesting research directions and open problems, which will be briefly discussed in this chapter.

## 5.1 The Kneser-Poulsen conjecture: a channel capacity point of view

In the early 1940s, it was thought to be impossible to send information at a positive rate with negligible probability of error. Shannon [95] surprised the communication theory community by proving that the probability of error could be made nearly zero for all communication rates below a fixed number, which is called the capacity of the channel. A channel is usually modeled as a transition kernel between two measurable spaces, and both the input $X$ and output $Y$ are assumed to be random. To define channel capacity, also known as Shannon capacity, a quantity called mutual information was introduced. For two jointly distributed random variables $X$, $Y$, say, in the discrete setting, the mutual information $I(X;Y)$ between $X$ and $Y$ can be defined as $I(X;Y) := H(X) - H(X|Y) = H(Y) - H(Y|X)$, where $H(X|Y)$ is called conditional entropy. The Shannon capacity is defined as $C := \sup_X I(X;Y)$. In order to have a parametric generalization of the Shannon capacity, we need a parametric generalization of the mutual information. Sibson [97] proposed one such parametric generalization as follows.

**Definition 5.1.1.** *Let $(A, \mathcal{A}, \nu)$ be a measure space, for any set of probability measures $\mathcal{W}$ absolutely continuous with respect to the $\sigma$-finite reference measure $\nu$ on the measure* *space, probability mass function $p$ on $\mathcal{W}$, the order $\alpha$ mean measure for the prior $p$ is defined as*

$$\frac{d\mu_{\alpha,p}}{d\nu} \triangleq \begin{cases} \prod_{w:p(w)>0} \left(\frac{dw}{d\nu}\right)^{p(w)} & \text{if } \alpha = 0, \\ \left(\sum_w p(w) \left(\frac{dw}{d\nu}\right)^{\alpha}\right)^{1/\alpha} & \text{if } \alpha \in \mathbb{R}_+, \\ \max_{w:p(w)>0} \frac{dw}{d\nu} & \text{if } \alpha = \infty. \end{cases}$$

*Then for order $\alpha \in (0,1) \cup (1,\infty)$, the Rényi mutual information for the prior $p$ is*

$$I_\alpha(p;\mathcal{W}) \triangleq \frac{\alpha}{\alpha-1} \log \|\mu_{\alpha,p}\|,$$

*where $\|\mu_{\alpha,p}\|$ denotes the total variation norm of $\mu_{\alpha,p}$.*

The Rényi mutual information for special orders $\alpha$ are obtained by taking limits. In particular, when $\alpha = \infty$, $I_\infty(p;\mathcal{W}) = \log \|\mu_{\infty,p}\|$. Let $\alpha$ be an order in $[0,\infty]$. The order $\alpha$ Rényi capacity of $\mathcal{W}$ ($\mathcal{W}$ can be thought of as a channel) is

$$C_{\alpha,\mathcal{W}} \triangleq \sup_{p\in\mathcal{P}(\mathcal{W})} I_\alpha(p;\mathcal{W}).$$

There are at least two other ways to define the Rényi mutual information: one by Arimoto [5] and another one by Augustin [7] and Csiszár [35]. A review of these three definitions of the Rényi mutual information has recently been provided by Verdú [101]. Despite the differences among these definitions, they all coincide with the Shannon mutual information when $\alpha = 1$. The equivalences and relationships between these definitions for Rényi mutual information are further extended in [3].

To see the connection to the Kneser-Poulsen conjecture, let $K = \{x_1, \ldots, x_m\}$ be a finite subset in $\mathbb{R}^d$, and $\mathcal{W}_K = \{x_i + \text{Uniform}(\mathcal{B}) : x_i \in K\}$ and let $p$ be a point mass function on $K$, which can be viewed as a point mass function on $\mathcal{W}$. Then one has $\frac{\mathrm{d}\mu_{\infty,p}}{\mathrm{d}x} = \max_{x_i:p(x_i)>0} \frac{\mathbf{1}_{\mathcal{B}(x_i)}(x)}{\mathrm{Vol}_d(\mathcal{B})}$. Thus,

$$I_\infty(p;\mathcal{W}_K) = \log \|\mu_{\infty,p}\| = \log \frac{\mathrm{Vol}_d(\bigcup_{x_i:p(x_i)>0} \mathcal{B}(x_i))}{\mathrm{Vol}_d(\mathcal{B})}.$$

Therefore, we have

$$C_{\infty,\mathcal{W}_K} = \sup_{p\in\mathcal{P}(\mathcal{W})} I_\infty(p;\mathcal{W}_K) = \log\frac{\mathrm{Vol}_d(K+\mathcal{B})}{\mathrm{Vol}_d(\mathcal{B})}.$$

Then the Kneser-Poulsen conjecture is equivalent to the following, for any contraction $T$,

$$C_{\infty,\mathcal{W}_K} \geq C_{\infty,\mathcal{W}_{T(K)}}.$$

We may think of Uniform($\mathcal{B}$) as a random noise when transmitting information, and if $x$ is sent, we will decode $y$ that we receive as "$x$", provided that $\|x-y\|_2 < 1$. If the pairwise distance between data in $K$ ($K$ can be thought of as the source data that will be sent) is larger, then the error probability should be smaller. Therefore, one may naturally believe that the communication rate (channel capacity) of the channel $\mathcal{W}_K$ should be larger than that of $\mathcal{W}_{T(K)}$. Built upon this intuition, we close this section with the following conjecture:

**Conjecture 10.** Let $K = \{x_1,\ldots,x_m\}$ be a finite subset in $\mathbb{R}^d$, $W$ be a radially symmetric log-concave random variable in $\mathbb{R}^d$, and let $\mathcal{W}_K = \{x_i + W : x_i \in K\}$. Then for $\alpha\in[0,\infty]$, and any contraction $T$,

$$C_{\alpha,\mathcal{W}_K} \geq C_{\alpha,\mathcal{W}_{T(K)}}.$$

### 5.2 The Kneser-Poulsen conjecture: a diversity point of view

Diversity was first introduced by Cobbold and Leinster [67] to measure the biodiversity of the ecosystem. Roughly speaking, diversity tells us the "effective number" of species in the ecosystem. In particular, for a finite metric space $(X,d)$ with the similarity kernel function $e^{-d(x,y)}$, diversity indicates the "effective number" of points in the metric space.

Meanwhile, for a finite set $K = \{x_1,\ldots,x_n\}\subset\mathbb{R}^d$, consider the quantity

$$\theta(K) := \frac{\mathrm{Vol}_d(K+r\mathcal{B})}{\mathrm{Vol}_d(r\mathcal{B})}.$$

Loosely speaking, this quantity tells us the "effective number" of centers in the set $K$. Since diversity of a finite metric space reveals the "effective number" of points in the metric space, one naturally wonders if we can connect $\theta(K)$ with diversity of the metric space $(K, d)$ for some suitably chosen metric $d$. While keeping the above in mind, we notice that

$$\delta(x,y) := \frac{\mathrm{Vol}_d(\{x,y\} + r\mathcal{B})}{\mathrm{Vol}_d(r\mathcal{B})} - 1 = \frac{1}{2\mathrm{Vol}_d(r\mathcal{B})} \int_{\mathbb{R}^d} |\mathbf{1}_{\mathcal{B}_r(x)}(z) - \mathbf{1}_{\mathcal{B}_r(y)}(z)| \, \mathrm{d}z$$

defines a metric on $\mathbb{R}^d$. In fact, the metric space $(\mathbb{R}^d, \delta)$ is isometrically embedded into $L^1(\mathbb{R}^d)$. We observe that this metric "preserves contraction". To be precise, for any contraction $T$ defined with respect to the Euclidean norm, one has

$$\delta(T(x), T(y)) \leq \delta(x, y).$$

This follows immediately from the fact that

$$\delta(x,y) = \frac{\mathrm{Vol}_d(\mathcal{B}_r(x) \triangle \mathcal{B}_r(y))}{2\mathrm{Vol}_d(r\mathcal{B})},$$

where $\mathcal{B}_r(x) \triangle \mathcal{B}_r(y)$ denotes the symmetric difference between the two sets.

Note that any probability measure $p$ defined on $K$ naturally induces a probability measure on $T(K)$ via the map $T$, i.e., $T\#p$ , the pushforward measure of $p$ by the map $T$. Since $T : (K, \delta) \mapsto (T(K), \delta)$ is a contraction, by the definition of diversity, one has the following inequality for any probability measure $p$ defined on $K$ and all orders $\alpha$.

$$D_\alpha^K(p) \geq D_\alpha^{T(K)}(T\#p).$$

From which, by recalling the definition of maximum diversity, we have

**Proposition 5.2.1.** *Assume that $\mathbb{R}^d$ is equipped with the metric $\delta$. Then for a finite* *subspace* $K = \{x_1, \ldots, x_n\}$ *and any contraction* $T : (\mathbb{R}^d, \|\cdot\|_2) \mapsto (\mathbb{R}^d, \|\cdot\|_2)$, *one has*

$$|K|_+ \geq |T(K)|_+.$$

If we are able to show that for any finite set $K$, $\theta(K)$ is an increasing function of $|K|_+$, then one proves the Kneser-Poulsen conjecture.

### 5.3 A log-concavity conjecture for diversity of order 1

Let $\Delta_n$=$\{p \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1,\ p_i \geq 0.\}$. It is a well known fact in information theory that $H_1(p) := -\sum_{i=1}^n p_i \log p_i$ is a concave function on $\Delta_n$ (see, e.g., [31]). Let $\mathcal{M}$ denote the following set:

$$\mathcal{M} := \{K : K \text{ positive definite matrix},\ K_{ii} = 1,\ 0 \leq K_{ij} \leq 1,\ 1 \leq i < j \leq n\}.$$

It was conjectured in [88] that for all $K \in \mathcal{M}$, the function $H_1^K : \Delta_n \mapsto \mathbb{R}$, defined by

$$H_1^K(p) := -\sum_{i=1}^n p_i \log(Kp)_i$$

is strictly concave on the interior of $\Delta_n$.

*Remark.* The concavity of $H_1^K(\cdot)$ implies the concavity of Shannon entropy. Indeed, let $\lambda = (\lambda_1, \ldots, \lambda_m) \in \Delta_m$, and for each $i$, $p_i(x)$ be a density function on $\mathbb{R}$. Then,

$$H_1^K\left(\sum_i \lambda_i p_i(x)\right) = -\int_{\mathbb{R}} \left(\sum_i \lambda_i p_i(x)\right) \log\left(\sum_i \lambda_i (K * p_i(x))\right)\, \mathrm{d}x$$

$$= -\int_{\mathbb{R}} \left(\sum_i \lambda_i p_i(x)\right) \log\left(\frac{\sum_i \lambda_i (K * p_i(x))}{\sum_i \lambda_i p_i(x)}\right)\, \mathrm{d}x + h\left(\sum_i \lambda_i p_i(x)\right)$$

By the convexity of the function $f(\alpha, x) := -\alpha \log(\frac{x}{\alpha})$ on $(0,\infty) \times (0,\infty)$, we have,

$$-\int_{\mathbb{R}} \left(\sum_i \lambda_i p_i(x)\right) \log\left(\frac{\sum_i \lambda_i (K * p_i(x))}{\sum_i \lambda_i p_i(x)}\right)\, \mathrm{d}x \leq -\int_{\mathbb{R}} \sum_i \lambda_i \left(p_i(x) \log \frac{K * p_i(x)}{p_i(x)}\right).$$

Hence, we have

$$H_1^K\left(\sum_i \lambda_i p_i(x)\right) - \sum_i \lambda_i H_1^K(p_i(x)) \leq h\left(\sum_i \lambda_i p_i(x)\right) - \sum_i \lambda_i h(p_i(x)),$$

as desired.

This conjecture was disproved by a counterexample kindly shared by Mark Meckes [78]. However, the counterexample does not rule out the possibility that the conjecture holds under stronger conditions on the matrix $K$. Specifically, we ask the following question.

**Question 3.** Let $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ be a finite metric space equipped with a metric $d$, such that the $n \times n$ matrix $(e^{-td(x_i,x_j)})_{i,j}$ is positive definite for all $t > 0$. Then for $K = (e^{-d(x_i,x_j)})_{i,j}$, is it true that $H_1^K(p)$ is strictly concave on the interior of $\Delta_n$?

Some computations related to this conjecture are presented in the Appendix.

## BIBLIOGRAPHY


[1] Gautam Aishwarya, Irfan Alam, Dongbin Li, Sergii Myroshnychenko, and Oscar Zatarain-Vera. Entropic exercises around the Kneser-Poulsen conjecture. *arXiv preprint arXiv:2210.12842*, 2022.

[2] Gautam Aishwarya, Dongbin Li, Mokshay Madiman, and Mark Meckes. Diversity and dimension of probability measures on metric spaces. *forthcoming paper*, 2022.

[3] Gautam Aishwarya and Mokshay Madiman. Conditional rényi entropy and the relationships between rényi capacities. *Entropy*, 22(5):526, 2020.

[4] Theodore W Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955.

[5] S Arimoto. Information measures and capacity of order $\alpha$ for discrete memoryless channels. *Topics in information theory*, 1977.

[6] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015.

[7] U. Augustin. *Noisy channels.* Univ. Erlangen-Nürnberg, Erlangen, West Germany, 1978.

[8] Keith Ball. Logarithmically concave functions and sections of convex sets in rn. *Studia Math*, 88(1):69–84, 1988.

[9] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[10] Károly Bezdek. *Lectures on sphere arrangements-the discrete geometric side.* Springer.

[11] Károly Bezdek and Robert Connelly. Pushing disks apart-the kneser-poulsen conjecture in the plane. 2002.

[12] Károly Bezdek and Márton Naszódi. The kneser–poulsen conjecture for special contractions. *Discrete & Computational Geometry*, 60(4):967–980, 2018.

[13] Maciej Białobrzeski and Piotr Nayar. Rényi entropy and variance comparison for symmetric log-concave random variables. *arXiv preprint arXiv:2108.10100*, 2021.

[14] S. Bobkov and M. Madiman. Concentration of the information in data with log-concave distributions. *Ann. Probab.*, 39(4):1528–1543, 2011.

[15] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, August 2011.

[16] S. Bobkov and M. Madiman. Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.*, 262:3309–3339, 2012.

[17] S. G. Bobkov. On concentration of distributions of random weighted sums. *Ann. Probab.*, 31(1):195–215, 2003.

[18] Sergey Bobkov and Mokshay Madiman. Entropy and the hyperplane conjecture in convex geometry. In *2010 IEEE International Symposium on Information Theory*, pages 1438–1442. IEEE, 2010.

[19] F. Bolley, I. Gentil, and A. Guillin. Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp-Lieb inequalities. *Ann. Probab.*, 46(1):261–301, 2018.

[20] C. Borell. Complements of Lyapunov's inequality. *Math. Ann.*, 205:323–331, 1973.

[21] H. J. Brascamp and E. H. Lieb. Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Advances in Math.*, 20(2):151–173, 1976.

[22] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. *Geometry of isotropic convex bodies*, volume 196 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2014.

[23] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[24] D. Bryant, K. T. Huber, V. Moulton, and P. F Tupper. Diversities and the generalized circumradius. *Preprint,* `arXiv:2110.13383`, 2021.

[25] D. Bryant and P. F. Tupper. Hyperconvexity and tight-span theory for diversities. *Adv. Math.*, 231(6):3172–3198, 2012.

[26] A. Burchard. A short course on rearrangement inequalities. *Available online at* `http://www.math.utoronto.ca/almut/rearrange.pdf`, June 2009.

[27] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000.

[28] Max Costa. A new entropy power inequality. *IEEE Transactions on Information Theory*, 31(6):751–760, 1985.

[29] M.H.M. Costa and T.M. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Trans. Inform. Theory*, 30(6):837–839, 1984.

[30] Thomas A Courtade, Guangyue Han, and Yaochen Wu. A counterexample to the vector generalization of costa's epi, and partial resolution. *arXiv preprint arXiv:1704.06164*, 2017.

[31] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[32] Balázs Csikós. On the volume of the union of balls. *Discrete & Computational Geometry*, 20(4):449–461, 1998.

[33] Balázs Csikós. On the volume of boolean expressions of balls–a review of the kneser–poulsen conjecture. *New Trends in Intuitive Geometry*, pages 65–94, 2018.

[34] Imre Csiszár. On the dimension and entropy of order $\alpha$ of the mixture of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 13(3-4):245–255, 1962.

[35] Imre Csiszár. Generalized cutoff rates and rényi's information measures. *IEEE Transactions on information theory*, 41(1):26–34, 1995.

[36] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer-Verlag, New York, second edition, 1998.

[37] Amir Dembo, Thomas M Cover, and Joy A Thomas. Information theoretic inequalities. *IEEE Transactions on Information theory*, 37(6):1501–1518, 1991.

[38] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[39] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL, 1992.

[40] W. Feller. *An Introduction to Probability Theory and its Applications. Vol. II.* John Wiley & Sons Inc., New York, second edition, 1971.

[41] M. Fradelizi, J. Li, and M. Madiman. Concentration of information content for convex measures. *Electron. J. Probab.*, 25(20):1–22, 2020. Available online at `arXiv:1512.01490v3`.

[42] M. Fradelizi, M. Madiman, M. Meyer, and A. Zvavitch. On the volume of the Minkowski sum of zonoids. *Preprint*, 2022.

[43] M. Fradelizi, M. Madiman, and L. Wang. Optimal concentration of information content for log-concave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret, and J. Rosinski, editors, *High Dimensional Probability VII: The Cargèse Volume*, Progress in Probability. Birkhäuser, Basel, 2016. Available online at `arXiv:1508.04093`.

[44] M. Fradelizi, M. Madiman, and A. Zvavitch. Sumset estimates in convex geometry. *Preprint*, 2022.

[45] M. Fradelizi and M. Meyer. Increasing functions and inverse Santaló inequality for unconditional functions. *Positivity*, 12(3):407–420, 2008.

[46] S. Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39:55–72, 1978.

[47] S. Fujishige. *Submodular functions and optimization*, volume 58 of *Annals of Discrete Mathematics*. Elsevier B. V., Amsterdam, second edition, 2005.

[48] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405 (electronic), 2002.

[49] R. J. Gardner and P. Gronchi. A Brunn-Minkowski inequality for the integer lattice. *Trans. Amer. Math. Soc.*, 353(10):3995–4024 (electronic), 2001.

[50] L. Grafakos. *Modern Fourier analysis*, volume 250 of *Graduate Texts in Mathematics*. Springer, New York, third edition, 2014.

[51] D. S. Grebenkov and B.-T. Nguyen. Geometrical structure of Laplacian eigenfunctions. *SIAM Rev.*, 55(4):601–667, 2013.

[52] Mikhael Gromov. Monotonicity of the volume of intersection of balls. In *Geometrical aspects of functional analysis*, pages 1–4. Springer, 1987.

[53] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[54] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[55] Mojzesz Kirszbraun. Über die zusammenziehende und lipschitzsche transformationen. *Fundamenta Mathematicae*, 22(1):77–108, 1934.

[56] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.

[57] B. Klartag. Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.*, 245(1):284–310, 2007.

[58] B. Klartag and V. D. Milman. Geometry of log-concave functions and measures. *Geom. Dedicata*, 112:169–182, 2005.

[59] Victor Klee and Stan Wagon. *Old and new unsolved problems in plane geometry and number theory.* Number 11. Cambridge University Press, 1991.

[60] Martin Kneser. Einige bemerkungen über das minkowskische flächenmass. *Archiv der Mathematik*, 6(5):382–390, 1955.

[61] Ioannis Kontoyiannis and Mokshay Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE transactions on information theory*, 60(8):4503–4514, 2014.

[62] L. Leindler. On a certain converse of Hölder's inequality. In *Linear operators and approximation (Proc. Conf., Oberwolfach, 1971)*, pages 182–184. Internat. Ser. Numer. Math., Vol. 20. Birkhäuser, Basel, 1972.

[63] T. Leinster. The magnitude of metric spaces. *Doc. Math.*, 18:857–905, 2013.

[64] T. Leinster and M. W Meckes. Maximizing diversity in biology and beyond. *Entropy*, 18(3):88, 2016.

[65] T. Leinster and M. W. Meckes. The magnitude of a metric space: from category theory to geometric measure theory. In *Measure theory in non-smooth spaces*, Partial Differ. Equ. Meas. Theory, pages 156–193. De Gruyter Open, Warsaw, 2017.

[66] T. Leinster and E. Roff. The maximum entropy of a metric space. *Q. J. Math.*, 72(4):1271–1309, 2021.

[67] Tom Leinster and Christina A Cobbold. Measuring diversity: the importance of species similarity. *Ecology*, 93(3):477–489, 2012.

[68] Dongbin Li and Mokshay Madiman. On a central limit theorem of information content for log-concave densities. *forthcoming paper*, 2022.

[69] E. H. Lieb. Proof of an entropy conjecture of Wehrl. *Comm. Math. Phys.*, 62(1):35–41, 1978.

[70] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, second edition, 2001.

[71] Ruoheng Liu, Tie Liu, H Vincent Poor, and Shlomo Shamai. A vector generalization of costa's entropy-power inequality with applications. *IEEE Transactions on Information Theory*, 56(4):1865–1879, 2010.

[72] M. Madiman and A.R. Barron. The monotonicity of information in the central limit theorem and entropy power inequalities. In *Proc. IEEE Intl. Symp. Inform. Theory*, pages 1021–1025. Seattle, July 2006.

[73] M. Madiman and I. Kontoyiannis. Entropy bounds on abelian groups and the Ruzsa divergence. *IEEE Trans. Inform. Theory*, 64(1):77–92, January 2018. Available online at `arXiv:1508.04089`.

[74] M. Madiman, A. Marcus, and P. Tetali. Entropy and set cardinality inequalities for partition-determined functions. *Random Struct. Alg.*, 40:399–424, 2012.

[75] M. Madiman, J. Melbourne, and P. Xu. Forward and reverse entropy power inequalities in convex geometry. In E. Carlen, M. Madiman, and E. M. Werner, editors, *Convexity and Concentration*, volume 161 of *IMA Volumes in Mathematics and its Applications*, pages 427–485. Springer, 2017. Available online at `arXiv:1604.04225`.

[76] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inform. Theory*, 56(6):2699–2713, June 2010.

[77] Mokshay Madiman, Liyao Wang, and Sergey Bobkov. Some applications of the nonasymptotic equipartition property of log-concave distributions. *Preprint*, 2016.

[78] Meckes Mark. private communication, 2022.

[79] M. W. Meckes. Positive definite metric spaces. *Positivity*, 17(3):733–757, 2013.

[80] M. W. Meckes. Magnitude, diversity, capacities, and dimensions of metric spaces. *Potential Anal.*, 42(2):549–572, 2015.

[81] James Melbourne. Rearrangement and prékopa–leindler type inequalities. In *High Dimensional Probability VIII*, pages 71–97. Springer, 2019.

[82] M. Merkle. Completely monotone functions: a digest. *Preprint*, 2012.

[83] V. D. Milman. Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(1):25–28, 1986.

[84] Hermann Minkowski. Theorie der konvexen körper, insbesondere begründung ihres oberflächenbegriffs. *Gesammelte Abhandlungen*, 2:131–229, 1911.

[85] V. H. Nguyen. *Inégalités fonctionnelles et convexité.* PhD thesis, Université Pierre et Marie Curie (Paris VI), October 2013.

[86] Grigoris Paouris and Peter Pivovarov. Intrinsic volumes and linear contractions. *Proceedings of the American Mathematical Society*, pages 1805–1808, 2013.

[87] G. Petridis. New proofs of Plünnecke-type estimates for product sets in groups. *Combinatorica*, 32(6):721–733, 2012.

[88] Jose Gallego Posada, Ankit Vani, Max Schwarzer, and Simon Lacoste-Julien. Gait: A geometric approach to information theory. In *International Conference on Artificial Intelligence and Statistics*, pages 2601–2611. PMLR, 2020.

[89] E. T. Poulsen. Problem 10. *Math. Scand., 2*, 1954.

[90] Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1):193–215, 1959.

[91] Alfréd Rényi et al. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1. Berkeley, California, USA, 1961.

[92] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[93] C. A. Rogers and G. C. Shephard. Some external problems for convex bodies. *Mathematika 5*, pages 93–102, 1958.

[94] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, expanded edition, 2014.

[95] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[96] Geoffrey C Shephard. Shadow systems of convex sets. *Israel Journal of Mathematics*, 2(4):229–236, 1964.

[97] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(2):149–160, 1969.

[98] A.J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

[99] T. Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010.

[100] Gábor Fejes Tóth and Włodzimierz Kuperberg. Four classic problems. *arXiv preprint arXiv:2202.09863*, 2022.

[101] Sergio Verdú. $\alpha$-mutual information. In *2015 Information Theory and Applications Workshop (ITA)*, pages 1–6. IEEE, 2015.

[102] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[103] L. Wang. *Heat capacity bound, energy fluctuations and convexity*. PhD thesis, Yale University, May 2014.

[104] Liyao Wang and Mokshay Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Transactions on Information Theory*, 60(9):5116–5137, 2014.

[105] William Henry Young. On the multiplication of successions of fourier constants. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 87(596):331–339, 1912.

[106] J. Zhang and R.W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory*, 44:1440–1452, 1998.

## Appendix

## COMPUTATIONS RELATED TO THE LOG-CONCAVITY CONJECTURE

Define a class of positive definite kernels as follows:

$$\mathcal{K} := \{K : K \text{ positive definite matrix},\ K_{ii} = 1,\ 0 \leq K_{ij} \leq 1,\ 1 \leq i < j \leq n\}.$$

Fix any $K \in \mathcal{K}$, and any $p \in \Delta_n$, where $\Delta_n \triangleq \{p \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1,\ p_i \geq 0\}$, define:

$$H_1^K(p) = -\sum_{i=1}^n p_i \log(Kp)_i.$$

It was conjectured in [88] that $H_1^K(p)$ is strictly concave in the interior of $\Delta_n$.

A direct calculation reveals that the Hessian of $-H_1^K(p)$ is

$$-\nabla_p^2[H_1^K(p)] = K\text{diag}\left(\frac{1}{Kp}\right) + \text{diag}\left(\frac{1}{Kp}\right)K - K\left(\text{diag}\frac{p}{(Kp)^2}\right)K,$$

where $\text{diag}\left(\frac{1}{Kp}\right)$ and $\left(\text{diag}\frac{p}{(Kp)^2}\right)$ represent the diagonal matrices with $i$-th diagonal entry being $\frac{1}{(Kp)_i}$ and $\frac{p_i}{(Kp)_i^2}$, respectively for $i = 1, \cdots, n$. It suffices to show the positive definiteness of $-\nabla_p^2[H_1^K(p)]$ in the interior of $\Delta_n$. Setting $A = K\text{diag}\left(\frac{1}{Kp}\right)$, one obtains,

$$-\nabla_p^2[H_1^K(p)] = A + A^T - A\ \text{diag}(p)\ A^T,$$

where $\text{diag}(p)$ is the diagonal matrix with $i$-th diagonal entry being $p_i$.

Let us assume that the probability vector $p$ is fully supported. Denote by $\Lambda$ the diagonal matrix $\mathrm{diag}(p)$. We have the following decomposition:

$$
\begin{aligned}
-\nabla_p^2[H_1^K(p)] &= A + A^T - A\Lambda A^T \\
&= \Lambda^{-1/2}\left[I_n - \left(\Lambda^{1/2}(I_n - A\Lambda)\Lambda^{-1/2}\right)\left(\Lambda^{-1/2}(I_n - \Lambda A^T)\Lambda^{1/2}\right)\right]\Lambda^{-1/2}
\end{aligned}
$$

Since $p$ is fully supported, the positiveness of $-\nabla_p^2[H_1^K(p)]$ is equivalent to the positiveness of

$$
C = I_n - \left(\Lambda^{1/2}(I_n - A\Lambda)\Lambda^{-1/2}\right)\left(\Lambda^{-1/2}(I_n - \Lambda A^T)\Lambda^{1/2}\right). \tag{A.1}
$$

To analyze the spectrum of $A\Lambda$, we do the following decomposition

$$
\begin{aligned}
A\Lambda &= K\mathrm{diag}\left(\frac{p}{Kp}\right) \\
&= \left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{-1/2}\left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{1/2} K \left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{1/2}\left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{1/2}.
\end{aligned}
$$

Therefore,

$$
A\Lambda \sim \left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{1/2} K \left(\mathrm{diag}\left(\frac{p}{Kp}\right)\right)^{1/2} \triangleq D
$$

which implies that all eigenvalues of $A\Lambda$ are real and positive:

$$
\lambda(A\Lambda) > 0.
$$

Note that $A\Lambda$ is a matrix with nonnegative entries, and the vector $q = ((Kp)_1, \cdots, (Kp)_n)^T$, the entries of which are strictly positive, is an eigenvector for $A\Lambda$ with corresponding eigenvalue being 1. Denote by $\sigma(M)$ the spectral radius of a matrix $M$. We claim that

$$
\sigma(A\Lambda) = 1.
$$

Indeed, since $A\Lambda$ is a matrix with nonnegative entries, by Perron-Frobenius theorem, for nonnegative matrix, there exists an eigenvector $q'$ with nonnegative entries, and the corresponding eigenvalue $\lambda_{q'}$ equals to $\sigma(A\Lambda)$. Meanwhile, by (3) and (4), we have that both 1 and $\lambda_{q'}$ are eigenvalues for $D$, with corresponding eigenvector $w = \left(\operatorname{diag}\left(\frac{p}{Kp}\right)\right)^{1/2} q$ and $w' = \left(\operatorname{diag}\left(\frac{p}{Kp}\right)\right)^{1/2} q'$, respectively. If $1 < \lambda_{q'}$, since $D$ is symmetric, it is necessary that

$$\langle w, w' \rangle = 0.$$

However, since $q$ has positive entries, $q'$ has nonnegative entries with at least one entry being non-zero, and $\left(\operatorname{diag}\left(\frac{p}{Kp}\right)\right)^{1/2}$ has positive entries, (7) can not hold. The claim now follows.

The key point now is to prove the positive definiteness of $C$ defined in (A.1). To simplify notation, let us set

$$\Lambda_2 = \operatorname{diag}\left(\frac{1}{Kp}\right),$$

$$N = \Lambda_2^{1/2} \left(I_n - D\right) \Lambda_2^{-1/2}.$$

To bring the symmetric matrix $D$ into the picture, let us rewrite (A.1) as follows:

$$C = \Lambda^{-1/2} \left[I_n - \left(\Lambda_2^{-1/2} \left(I_n - D\right) \Lambda_2^{1/2}\right) \left(\Lambda_2^{1/2} \left(I_n - D\right) \Lambda_2^{-1/2}\right)\right] \Lambda^{-1/2}.$$

$$= \Lambda^{-1/2} \left[I_n - N^T N\right] \Lambda^{-1/2} \tag{A.2}$$

Let $Q$ be the orthogonal matrix such that

$$Q^T (I_n - D) Q = \Lambda_1,$$

where $\Lambda_1$ is a diagonal matrix. Note that $D$ is positive definite, and we have shown that its largest eigenvalue is 1 with multiplicity 1. Therefore $I_n - D$ is a singular $M$ matrix and $\Lambda_1$ is singular with one diagonal entry being 0.

To prove concavity, by (A.2), we need to show the operator norm of $N$ is strictly less than 1, i.e.,

$$\|N\| < 1.$$

Since $\Lambda_2^{1/2}Q$ is non-singular, the columns of it are linearly independent. Therefore, any unit vector $u$ can be expressed as

$$u = \Lambda_2^{1/2}Q\alpha$$

for some vector $\alpha \in \mathbb{R}^n$. From this, we obtain

$$\|N\| = \sup_{\|u\|_2=1} \|Nu\|_2 = \sup_{\|\Lambda_2^{1/2}Q\alpha\|_2=1} \|\Lambda_2^{1/2}Q\Lambda_1\alpha\|_2. \tag{A.3}$$

Define an ellipsoid as follows:

$$\mathcal{E} := \{\alpha : \|\Lambda_2^{1/2}Q\alpha\|_2 \leq 1\}.$$

From (A.3), to prove $\|N\| < 1$, it is equivalent to show that

$$\Lambda_1\mathcal{E} \subset \mathcal{E}.$$

In other words, will the contraction $\Lambda_1$ under the usual Euclidean norm still be a contraction under the Euclidean norm induced by the ellipsoid $\mathcal{E}$? To answer the question, one may need to get the kernel matrix $K$ involved to deduce more structure about the ellipsoid $\mathcal{E}$...

Some other information that might be useful. Note that

$$h_{\mathcal{E}}(Q^T e_j) = \sqrt{(Kp)_j},$$

where $h_{\mathcal{E}}(\cdot)$ is the support function of $\mathcal{E}$, and $e_j$ is the unit vector with the $j-$th coordinate being 1. Whence, we have that $Q(\mathcal{E})$ is a standard ellipsoid, whose semi-axis in direction $e_i$ has length $\sqrt{(Kp)_i}$. We can also compare the width of two ellipsoid in the direction of the axes of $\mathcal{E}$:

$$h_{\mathcal{E}}(Q^T e_j) > h_{\Lambda_1 \mathcal{E}}(Q^T e_j), \quad j = 1, \cdots, n.$$

In other words, one can show that

$$\sup_{\alpha \in \partial \mathcal{E}} \langle \Lambda_1 \alpha, Q^T e_j \rangle < \sqrt{(Kp)_j}, \quad j = 1, \cdots, n,$$

which is equivalent to show

$$\| \Lambda_2^{-1/2} Q \Lambda_1 Q^T e_j \|_2 < \sqrt{(Kp)_j}.$$

To this end, note that

$$\| \Lambda_2^{-1/2} Q \Lambda_1 Q^T e_j \|_2^2 = \langle \Lambda_2^{-1} (I_n - D) e_j, (I_n - D) e_j \rangle.$$

Since $(I_n - D)e_j$ is the $j-$th column of the matrix $(I_n - D)$, we have

$$\langle \Lambda_2^{-1} (I_n - D) e_j, (I_n - D) e_j \rangle = (Kp)_j (1 - \frac{p_j}{(Kp)_j})^2 + \sum_{l \neq j} \frac{p_j K_{jl}^2 p_l}{(Kp)_j}.$$

$$< (Kp)_j (1 - \frac{p_j}{(Kp)_j})^2 + \sum_{l \neq j} \frac{p_j K_{jl} p_l}{(Kp)_j} = (Kp)_j (1 - \frac{p_j}{(Kp)_j})^2 + p_j (1 - \frac{p_j}{(Kp)_j}).$$

$$= (Kp)_j - p_j < (Kp)_j,$$

as desired.