Common Information, Noise Stability, and Their Extensions

Other titles in Foundations and Trends® in Communications and Information Theory

Asymptotic Frame Theory for Analog Coding
Marina Haikin, Matan Gavish, Dustin G. Mixon and Ram Zamir
ISBN: 978-1-68083-908-1

Modeling and Optimization of Latency in Erasure-coded Storage Systems
Vaneet Aggarwal and Tian Lan
ISBN: 978-1-68083-842-8

An Algebraic and Probabilistic Framework for Network Information Theory
S. Sandeep Pradhan, Arun Padakandla and Farhad Shirani
ISBN: 978-1-68083-766-7

Theoretical Foundations of Adversarial Binary Detection
Mauro Barni and Benedetta Tondi
ISBN: 978-1-68083-764-3

Polynomial Methods in Statistical Inference
Yihong Wu and Pengkun Yang
ISBN: 978-1-68083-730-8

Information-Theoretic Foundations of Mismatched Decoding
Jonathan Scarlett, Albert Guillen i Fabregas, Anelia Somekh-Baruch and Alfonso Martinez
ISBN: 978-1-68083-712-4

Common Information, Noise Stability, and Their Extensions

Lei Yu
Nankai University
China
leiyu@nankai.edu.cn

Vincent Y. F. Tan
National University of Singapore
Singapore
vtan@nus.edu.sg

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

L. Yu and V. Y. F. Tan. *Common Information, Noise Stability, and Their Extensions*. Foundations and Trends® in Communications and Information Theory, vol. 19, no. 2, pp. 107–389, 2022.

ISBN: 978-1-63828-015-6

Foundations and Trends® in Communications and Information Theory
Volume 19, Issue 2, 2022

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2022, Volume 19, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

Contents

Common Information, Noise Stability, and Their Extensions

Lei Yu[1] and Vincent Y. F. Tan[2]

[1] *School of Statistics and Data Science, LPMC, KLMDASR, and LEBPS, Nankai University, China; leiyu@nankai.edu.cn*
[2] *Department of Mathematics, Department of Electrical and Computer Engineering, Institute of Operations Research and Analytics, National University of Singapore, Singapore; vtan@nus.edu.sg*

ABSTRACT

Common information is ubiquitous in information theory and related areas such as theoretical computer science and discrete probability. However, because there are multiple notions of common information, a unified understanding of the deep interconnections between them is lacking. This monograph seeks to fill this gap by leveraging a small set of mathematical techniques that are applicable across seemingly disparate problems.

In Part I, we review the operational tasks and properties associated with Wyner's and Gács–Körner–Witsenhausen's (GKW's) common information. In Part II, we discuss extensions of the former from the perspective of distributed source simulation. This includes the Rényi common information which forms a bridge between Wyner's common information and the exact common information. Via a surprising equivalence between the Rényi common information of order ∞ and the exact common information, we demonstrate the existence of a joint source in which the exact common

---

Lei Yu and Vincent Y. F. Tan (2022), "Common Information, Noise Stability, and Their Extensions", Foundations and Trends® in Communications and Information Theory: Vol. 19, No. 2, pp 107–389. DOI: 10.1561/0100000122.

information strictly exceeds Wyner's common information. Other closely related topics discussed in Part II include the channel synthesis problem and the connection of Wyner's and exact common information to the nonnegative rank of matrices.

In Part III, recognizing that GKW's common information is zero for most non-degenerate sources, we examine it with a more refined lens via the Non-Interactive Correlation Distillation (NICD) problem in which we quantify the agreement probability of extracted bits from a bivariate source. We extend this to the noise stability problem which includes as special cases the k-user NICD and q-stability problems. This allows us to seamlessly transition to discussing their connections to various conjectures in information theory and discrete probability, such as the Courtade–Kumar, Li–Médard and Mossell–O'Donnell conjectures. Finally, we consider functional inequalities (e.g., the hypercontractivity and Brascamp–Lieb inequalities), which constitute a further generalization of the noise stability problem in which the Boolean functions therein are replaced by nonnegative functions. We demonstrate that the key ideas behind the proofs in Part III can be presented in a pedagogically coherent manner and unified via information-theoretic and Fourier-analytic methods.

1

Introduction

1.1 Motivation

Let X be the statistical description of a set of images whose foregrounds and backgrounds are those of an airplane and the blue sky respectively. Let Y, which is correlated to X, be the statistical description of another set of images whose foregrounds are those of a unicorn and the blue sky respectively. It seems natural and intuitive that the common information in X and Y should be the number of bits needed to describe the blue sky, which is the common part of X and Y. Can we make this observation precise and quantitative for *arbitrary* (X, Y) pairs? This monograph is centered on this fundamental question in information and probability theory. In other words, we would like to quantify, via an assortment of well-motivated measures, the *intrinsic similarity* or *common information* between two correlated random variables X and Y. Regardless of what applications there may be, the pursuit of operationally meaningful measures that quantify the common information between two random variables seems to be an extremely worthy academic endeavor. This is especially so for researchers in information and coding theory, theoretical computer science, and cryptography who are seeking to understand the inherent difficulties in generating correlated bits from a single joint source, or simulating a joint source using a single source of randomness in a distributed manner.

In probability, statistics, and data analysis, there are numerous popular functionals of joint distributions that quantify the amount of correlation or dependence between two random variables X and Y. If these random variables have joint distribution π_{XY} and means μ_X and μ_Y respectively, such paradigmatic examples include the *Pearson correlation coefficient*

$$\rho(X;Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{\mathbb{E}[(X-\mu_X)^2]\mathbb{E}[(Y-\mu_Y)^2]}} \tag{1.1}$$

and the *Hirschfeld–Gebelein–Rényi (HGR) maximal correlation*

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)), \tag{1.2}$$

where the supremum is taken over all real-valued functions f and g such that $0 < \mathrm{Var}(f(X)), \mathrm{Var}(g(Y)) < \infty$. In addition, an information-theoretic quantity known as the *mutual information*

$$I(X;Y) = \begin{cases} \displaystyle\int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{\mathrm{d}\pi_{XY}}{\mathrm{d}(\pi_X\pi_Y)}\right)\mathrm{d}\pi_{XY} & \text{if } \pi_{XY} \ll \pi_X\pi_Y \\ +\infty & \text{otherwise} \end{cases},$$

also serves to quantify the dependence between two random variables. These measures have the property that they are zero if the two random variables are independent, fulfilling a basic requirement of any measure that quantifies the dependence between two random variables. These measures can be regarded as common information quantities between X and Y, jointly distributed as π_{XY}. Indeed, the mutual information $I(X;Y)$ captures the amount of information about X provided by observing Y, as can observed in the celebrated distributed lossless compression theorem of Slepian and Wolf [41], [156]. Are there any other *operationally-motivated* measures that allow us to gain deeper insights on the common information between X and Y given their numerical values?

In information and coding theory, there are two canonical examples of operationally-motivated common information measures that have been widely accepted since their inceptions in the 1970s. The first, which was introduced in 1973, is *Gács–Körner–Witsenhausen's (GKW's) common information* [60], [178], defined as

$$C_{\mathrm{GKW}}(\pi_{XY}) := \sup_{f,g} H(f(X)), \tag{1.3}$$

where the supremum is taken over all pairs of deterministic functions (f, g) defined respectively on $\mathcal{X}$ and $\mathcal{Y}$ such that $f(X) = g(Y)$ with π_{XY}-probability one. The second, which was introduced in 1975, is *Wyner's common information* [182], defined as

$$C_{\mathrm{W}}(\pi_{XY}) := \inf_{P_W P_{X|W} P_{Y|W} : P_{XY} = \pi_{XY}} I_P(W; XY), \tag{1.4}$$

where the infimum extends over triples of random variables $(W, X, Y) \sim P_{WXY}$ such that $X - W - Y$ forms a Markov chain and $P_{XY} = \pi_{XY}$.

1.2 Overview of the Monograph

Our twin objectives in this monograph are as follows. Firstly, we seek to provide a concise review of these classical notions of common information. Secondly, we endeavor to connect these quantities to new notions of common information in the literature that have gained traction recently. A flowchart of the sections in this monograph is provided in Fig. 1.1.

1.2.1 Part I: Classic Common Information Quantities

We commence in Part I by reviewing the operational tasks associated with the classical common information quantities in (1.3) and (1.4) and describing their salient properties. This part consists of Sections 2 and 3 on Wyner's and GKW's common information respectively.

1.2.2 Part II: Extensions of Wyner's Common Information

We then extend and generalize Wyner's common information in Part II of this monograph, which consists of four sections. In Section 4, we review the *Rényi common information*, originally studied by the present authors [197], [202]. In his seminal paper [182], Wyner used the normalized relative entropy

$$\frac{1}{n} D\left(P_{X^nY^n} \| \pi_{XY}^n\right) = \frac{1}{n} \sum_{x^n, y^n} P_{X^nY^n}(x^n, y^n) \log \frac{P_{X^nY^n}(x^n, y^n)}{\pi_{XY}^n(x^n, y^n)}.$$

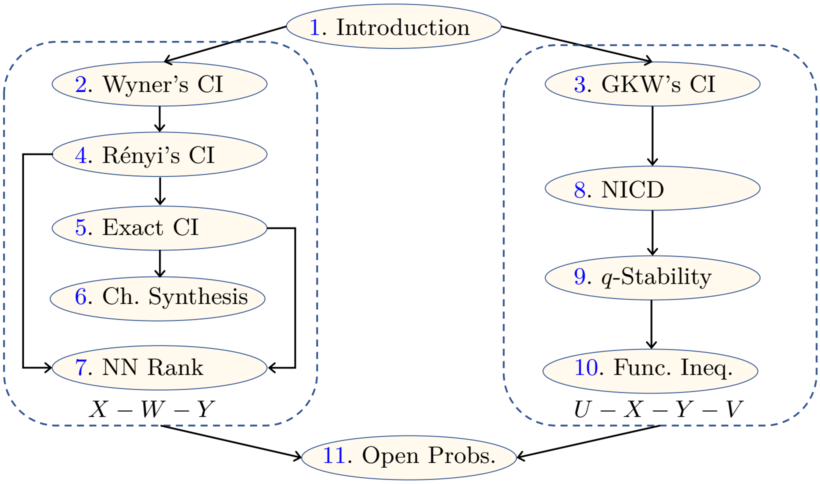

Figure 1.1: Flowchart of the sections in this monograph (CI, NN, and NICD stand for Common Information, Nonnegative, and Non-Interactive Correlation Distillation respectively)

to quantify the discrepancy between the synthesized distribution $P_{X^nY^n}$ and the target distribution π_{XY}^n and sought the minimum rate for distributed source synthesis for which this quantity vanishes as the blocklength n grows. The Rényi common information [197], [202] generalizes this to the case in which the discrepancy measures used belong to the families of normalized and unnormalized Rényi divergences. For Rényi order $1+s \in (0,\infty)\setminus\{1\}$, the unnormalized form can be expressed as

$$D_{1+s}(P_{X^nY^n}\|\pi_{XY}^n) = \frac{1}{s}\log\sum_{x^n,y^n} P_{X^nY^n}(x^n,y^n)\left(\frac{P_{X^nY^n}(x^n,y^n)}{\pi_{XY}^n(x^n,y^n)}\right)^s.$$

We use this family of measures to build a bridge to the topic of discussion in Section 5, namely, the *exact common information*, a quantity first defined and studied by Kumar, Li, and El Gamal [103]; see Definition 5.3 for its precise definition. In contrast to the Rényi common information, the exact version requires that synthesized distribution be *exactly* equal to the target distribution for some blocklength n; however, *variable-length* codes are permitted. Using an unexpected equivalence between the unnormalized Rényi common information of order ∞ (the limit of D_{1+s} as $s \to \infty$)

$$D_\infty(P_{X^nY^n} \| \pi_{XY}^n) = \log \max_{(x^n,y^n): P_{X^nY^n}(x^n,y^n)>0} \frac{P_{X^nY^n}(x^n, y^n)}{\pi_{XY}^n(x^n, y^n)},$$

and the exact common information, we argue that the latter can be strictly larger than Wyner's common information for some sources, specifically the doubly symmetric binary source (DSBS).

In Section 6, we use the preceding notions to describe the problem of *channel synthesis*. We review this problem in both the approximate and exact settings and show that it produces a continuum of common information measures that interpolate from the mutual information to Wyner's or exact common information.

In Section 7, we describe a seemingly tangential topic in numerical linear algebra, namely the *nonnegative rank* of nonnegative matrices [65], [168]. It turns out that this area of research has intimate connections to the preceding notions of common information, leading to some interesting open problems.

1.2.3 Part III: Extensions of Gács–Körner–Witsenhausen's Common Information

It is known that GKW's common information is zero for most non-pathological sources such as the doubly symmetric binary source and the bivariate Gaussian source. Consequently, in itself, GKW's common information does not provide any tangible quantification of how "similar" two sources are. The goal of Part III, which consists of three sections, is thus to consider several *refinements* of GKW's common information in which new insights can be readily gleaned.

We start in Section 8 by providing an extensive discussion of the 2-*user Non-Interactive Correlation Distillation* (NICD) problem [94], [124]. Given a pair of random vectors $(X^n, Y^n) \sim \pi_{XY}^n$ in which each (X_i, Y_i) is drawn independently from a DSBS, this problem concerns the agreement probability of the random bits that can be extracted from X^n and Y^n individually. In other words, we wish to quantify

$$\max_{f,g} \Pr(U = V) \quad \text{and} \quad \min_{f,g} \Pr(U = V),$$

where $U = f(X^n)$ and $V = g(Y^n)$ and f and g are $\{0,1\}$-valued (i.e., Boolean) functions such that the marginals $\Pr(U = 1)$ and $\Pr(V = 1)$ are appropriately constrained. For example, for the maximization version of the NICD problem, we place upper bounds on $\Pr(U = 1)$ and $\Pr(V = 1)$. We quantify these agreement probabilities by studying various geometric structures such as Hamming subcubes and Hamming balls. We discuss their optimality in several asymptotic regimes (such as the central limit or large deviations regimes) using results from concentration of measure and Boolean Fourier analysis, among other techniques.

In Section 9, we extend the NICD problem to the *multi-user* version. For the k-user case, there are k correlated sources $X_1^n, X_2^n, \ldots, X_k^n$ that are generated independently conditioned on another source Y^n such that the joint distribution of X_i^n and Y^n is π_{XY}^n. We are interested in quantifying

$$\max_{f_1, f_2, \ldots, f_k} \Pr\left(U_1 = U_2 = \ldots = U_k\right),$$

where $U_i = f_i(X_i^n), i = 1, 2, \ldots, k$ and the maximum extends over all k-tuples of Boolean functions f_i's whose marginals are also constrained by placing upper bounds on $\Pr(U_i = 1)$. We also discuss the connection of the k-user NICD problem to *q-stability* [52], [110] in which the number of users k is replaced by an arbitrary real number q. This allows us to seamlessly segue into a review of recent advances in contemporary conjectures in information theory and discrete probability. These include the Courtade–Kumar conjecture [40], the Mossel–O'Donnell conjecture [122], and the Li–Médard conjecture [110]. Mathematical tools used here include the analysis of Boolean functions [131] and, in particular, edge-isoperimetric inequalities and the study of the maximal degree-1 Fourier weight.

In Section 10, we connect these notions and results to *functional inequalities* including the hypercontractivity, the logarithmic Sobolev, the Brascamp–Lieb inequalities, as well as their strengthened counterparts. This section generalizes the preceding two sections in that the Boolean functions f_i are replaced by arbitrary nonnegative functions.

The monograph is concluded in Section 11 in which we summarize open problems in this fascinating area of study.

The common theme in Part II is the Markov chain $X - W - Y$; this corresponds to the constraint that defines Wyner's common information in (1.4). In contrast, in Part III, we focus on the Markov chain $U - X - Y - V$; this corresponds to the Markov chain in the NICD problem in which $U = f(X^n)$ and $V = g(Y^n)$ for some Boolean functions f and g. It is also present in GKW's common information. At first glance, this appears to be different from the constraint in (1.3). However, this constraint is merely a special case of $U - X - Y - V$ by taking U and V to be deterministic functions of X and Y respectively such that they are also constrained to be equal almost surely.

1.3 Notation

To appreciate the material in this monograph, the reader is expected to have some background in information theory at the level of Cover and Thomas [42]. We will also make frequent use of the method of types, for which an excellent exposition can be found in Csiszár and Körner [45].

In this monograph, we generally follow the notation in Cover and Thomas [42], El Gamal and Kim [51], and Csiszár and Körner [45].

1.3.1 Random Variables and Probability Distributions

Random variables and their realizations are denoted by upper case letters (such as X and Y) and lower case letters (such as x and y) respectively. The sets of values that the realizations take on, also called *alphabets*, are denoted by calligraphic letters such as $\mathcal{X}$ and $\mathcal{Y}$. We use $P_X, \tilde{P}_X, Q_X$, and π_X to denote various probability distributions on alphabet $\mathcal{X}$. If a random variable X is distributed according to P_X, we write $X \sim P_X$. As we work with both discrete and continuous random variables in this monograph, we will often have to distinguish between probability mass functions (PMFs) for discrete random variables and probability density functions (PDFs) for continuous random variables. If X is discrete, we use $x \in \mathcal{X} \mapsto P_X(x)$ to denote its PMF. The PDF of a (real-valued) continuous random variable is denoted as $f_X : x \in \mathbb{R} \mapsto (\mathrm{d}P_X/\mathrm{d}\mu)(x)$, where μ is the Lebesgue measure on $\mathbb{R}$. These will also be denoted as P or f when the random variable X is clear from the context. Throughout the monograph, the notations π_X and π_{XY} are reserved for *target* and *source distributions*.

The set of PMFs on $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$ and the set of conditional PMFs on $\mathcal{Y}$ given a variable taking values in $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{Y}|\mathcal{X}) = \{P_{Y|X} : P_{Y|X}(\cdot|x) \in \mathcal{P}(\mathcal{Y}), x \in \mathcal{X}\}$. The joint distribution induced by $P_X \in \mathcal{P}(\mathcal{X})$ and $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ is denoted as $P_X P_{Y|X} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The *support* of a discrete distribution is denoted as $\mathrm{supp}(P) := \{x \in \mathcal{X} : P(x) > 0\}$. Given an input distribution $P_X \in \mathcal{P}(\mathcal{X})$ and a conditional distribution $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, if the induced output distribution is $P_Y(y) = \sum_x P_X(x) P_{Y|X}(y|x)$ (for the discrete case), we write this as $P_X \to P_{Y|X} \to P_Y$. For two distributions P and Q (defined on the same measurable space), we use $P \ll Q$ to denote that P is *absolutely continuous* with respect to Q. In the finite alphabet case, $P \ll Q$ means that for every $x \in \mathcal{X}$ such that $Q(x) = 0$, it holds that $P(x) = 0$.

We say that three random variables X, Y, and Z *form a Markov chain in this order* if X and Z are conditionally independent given Y. In this case, we write $X - Y - Z$. For discrete random variables, $X - Y - Z$ if and only if $P_{XYZ}(x, y, z) = P_Y(y) P_{X|Y}(x|y) P_{Z|Y}(z|y)$ for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. As is customary in information theory, for two integers m and n, we write X_m^n to mean the random vector $(X_m, X_{m+1}, \ldots, X_n)$; when $m = 1$, this is abbreviated to X^n. A particular realization of X^n, a deterministic vector, is denoted as $x^n = (x_1, x_2, \ldots, x_n)$. We denote the *n-fold product distribution* of P as P^n, which is defined by the formula $P^n(x^n) = \prod_{i=1}^n P(x_i)$ for all $x^n \in \mathcal{X}^n$.

A *stationary memoryless source*, denoted by $X \sim P_X \in \mathcal{P}(\mathcal{X})$, is a discrete-time stochastic process $\{X_i\}_{i \in \mathbb{N}}$ such that X_i's are independent copies of X. We also denote a source X by its distribution P_X. We use X^n to denote the first n random variables in the stochastic process $\{X_i\}_{i \in \mathbb{N}}$. With a slight abuse of terminology, X^n is also called a *source sequence* of the source X. A *stationary memoryless channel*, denoted by $P_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, is a random transformation that outputs a length-n random vector $Y^n \sim P_{Y|X}^n(\cdot|x^n)$ if the input is the length-n vector $x^n \in \mathcal{X}^n$. Since we deal almost exclusively with stationary memoryless sources and channels in this monograph, we will omit the term "stationary memoryless" when we mention sources and channels.

We will work mainly with three types of random variables in this monograph. A discrete *uniform* random variable X takes equal probabilities on its support $\mathcal{X}$ and its probability distribution is denoted as $\mathrm{Unif}(\mathcal{X})$. A *Bernoulli* random variable X is one with support $\{0,1\}$. Its probability distribution is abbreviated as $\mathrm{Bern}(a)$ if $\Pr(X=1)=a$. A (d-dimensional) *normal* or *Gaussian* random variable or vector X has a PDF that is denoted by

$$\mathbf{x}\in\mathbb{R}^d \mapsto \mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d\,\det(\boldsymbol{\Sigma})}}\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big),$$

(or simply $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$) where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix respectively.

1.3.2 Types or Empirical Distributions

We will often use the method of types [45] in our calculations, especially for finite alphabets. Given a sequence $x^n\in\mathcal{X}^n$, we use

$$T_{x^n}(a) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{x_i=a\} \quad \text{for all } a\in\mathcal{X}$$

to denote its *type* or *empirical distribution*. The type of a length-n sequence will be denoted by T or $T_X^{(n)}$ depending on the context. The set of all sequences with type T is denoted as $\mathcal{T}_T\subset\mathcal{X}^n$. This is known as the *type class* of T. The set of all types that can be formed from sequences of length n taking values in alphabet $\mathcal{X}^n$ is denoted as $\mathcal{P}_n(\mathcal{X})$, which is a subset of the probability simplex $\mathcal{P}(\mathcal{X})$.

1.3.3 Information Measures

We now recap the necessary information measures used in this monograph. For $X\sim P_X$, we denote its *Shannon entropy* as

$$H(X)=H_P(X)=H(P_X) := -\sum_{x\in\mathrm{supp}(P_X)} P_X(x)\log P_X(x). \quad (1.5)$$

All logarithms are to the base 2 unless otherwise specified. For $(X, Y) \sim P_{XY}$, we denote the *conditional entropy* of X given Y as

$$\begin{aligned} H(X|Y) = H_P(X|Y) &= H(P_{X|Y}|P_Y) \\ &:= -\sum_{y\in\mathcal{Y}} P_Y(x) \sum_{x\in\mathrm{supp}(P_{X|Y}(\cdot|y))} P_{X|Y}(x|y)\log P_{X|Y}(x|y). \end{aligned}$$

The *mutual information* between X and Y where $(X, Y) \sim P_{XY}$ is denoted as

$$I_P(X;Y) = I(P_X, P_{Y|X}) := H_P(X) - H_P(X|Y).$$

The subscripts in H_P and I_P are used to emphasize the distribution of (X, Y) under which these information measures are computed. When the distribution is clear from the context, the subscripts will be omitted. The *relative entropy* or *Kullback–Leibler divergence* between two distributions P_X and Q_X defined on the same (countable) alphabet is[1]

$$D(P_X\|Q_X) := \sum_{x\in\mathrm{supp}(P_X)} P_X(x)\log\frac{P_X(x)}{Q_X(x)}.$$

The *conditional relative entropy* of two conditional distributions $P_{Y|X}$ and $Q_{Y|X}$, given a distribution P_X, is

$$D(P_{Y|X}\|Q_{Y|X}|P_X) := D(P_X P_{Y|X}\|P_X Q_{Y|X}). \tag{1.6}$$

In addition to the Shannon information measures above, we need to recap the family of Rényi information measures [144], [166] as this is central to the majority of our discussion in this monograph. For two distributions $P_X, Q_X \in \mathcal{P}(\mathcal{X})$ on a countable set $\mathcal{X}$, the *Rényi divergence* of order $1 + s \in (0, 1) \cup (1, \infty)$ is

$$D_{1+s}(P_X\|Q_X) := \frac{1}{s}\log\sum_{x\in\mathrm{supp}(P_X)} P_X(x)\left(\frac{P_X(x)}{Q_X(x)}\right)^s.$$

[1]This definition is only applicable when the alphabets are countable. For P_X and Q_X defined on a general probability space, the ratio P_X/Q_X should be replaced with the Radon–Nikodym derivative $\mathrm{d}P_X/\mathrm{d}Q_X$ (if $P_X \ll Q_X$), and the expectation with respect to P_X should be written as a Lebesgue integral over $\mathcal{X}$. If P_X is not absolutely continuous with respect to Q_X, $D(P_X\|Q_X)$ is defined to be $+\infty$. In the following, for simplicity, we only provide definitions of information-theoretic quantities for countable alphabets.

The Rényi divergence is monotonically nondecreasing in its order. Sibson's [155] version of the *conditional Rényi divergence* between two conditional distributions $P_{Y|X}$ and $Q_{Y|X}$ given a distribution P_X is

$$D_{1+s}(P_{Y|X}\|Q_{Y|X}|P_X) := D_{1+s}(P_X P_{Y|X}\|P_X Q_{Y|X}). \tag{1.7}$$

We note that while the conditional relative entropy in (1.6) is the expectation of $D(P_{Y|X}(\cdot|X)\|Q_{Y|X}(\cdot|X))$ over $X \sim P_X$, the conditional Rényi divergence in (1.7) depends on $D_{1+s}(P_{Y|X}(\cdot|X)\|Q_{Y|X}(\cdot|X))$ in a more involved way; indeed, it is a generalized mean of the random variable $D_{1+s}(P_{Y|X}(\cdot|X)\|Q_{Y|X}(\cdot|X))$ evaluated at s. For a more detailed discussion on this point, the reader is referred to Cai and Verdú [32]. We also note that there are other definitions of the conditional Rényi divergence but we will use the definition in (1.7) in this monograph; see [20], [43], [155]. The Rényi divergence and its conditional version in (1.7) can be extended to all orders $1+s \in \{0,1,\infty\}$ by taking the appropriate limits. In particular, when $s \to 0$, we recover the usual relative entropy. An order of the Rényi divergence that will be of particular interest to us in this monograph is the *Rényi divergence of order* ∞. This is the divergence we obtain when we let $s \to \infty$, i.e.,

$$D_\infty(P_X\|Q_X) := \log \sup_{x\in\mathrm{supp}(P_X)} \frac{P_X(x)}{Q_X(x)}.$$

The *Rényi entropy* of order $1+s \in (0,1)\cup(1,\infty)$ of a probability mass function $P_X \in \mathcal{P}(\mathcal{X})$ is defined as

$$H_{1+s}(P_X) = -\frac{1}{s}\log \sum_{x\in\mathrm{supp}(P_X)} \big(P_X(x)\big)^{1+s}. \tag{1.8}$$

It is easy to check that

$$H_{1+s}(P_X) := \log|\mathcal{X}| - D_{1+s}(P_X\|\mathrm{Unif}(\mathcal{X})). \tag{1.9}$$

Similarly to the Rényi divergence, we define $H_0(P_X)$ and $H_\infty(P_X)$ as the limits of $H_{1+s}(P_X)$ as $s \downarrow -1$ and $s \to \infty$ respectively. These are known as the *max-entropy* and *min-entropy* respectively. Of special importance is the case when $s \to 0$, in which case $H_{1+s}(P_X)$ reduces to the Shannon entropy defined in (1.5). Since the relation in (1.9) holds and the Rényi divergence is nondecreasing in its order, the Rényi entropy is nonincreasing in its order.

We need one additional measure of the discrepancy between two distributions. The *total variation distance* or simply the *TV distance* is defined for two distributions P and Q on a common (countable) alphabet $\mathcal{X}$ as

$$|P-Q| := \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x)-Q(x)|.$$

More generally, $|P-Q| = \sup_{\mathcal{A}\subset\mathcal{X}} |P(\mathcal{A})-Q(\mathcal{A})|$, where $\mathcal{A}$ runs over all (measurable) subsets of $\mathcal{X}$. Pinsker's inequality yields the following bound on the TV distance in terms of the relative entropy

$$|P-Q|^2 \le \frac{\ln 2}{2}\cdot D(P\|Q). \tag{1.10}$$

1.3.4 Typical Sets

In our achievability proofs, we will often need to use the notion of *typical sets* [42], [51], [135]. The *ϵ-strongly typical set* with respect to a distribution $P_X \in \mathcal{P}(\mathcal{X})$ is defined as

$$\mathcal{T}_\epsilon^{(n)}(P_X) := \Big\{x^n \in \mathcal{X}^n : |T_{x^n}(x) - P_X(x)| \le \epsilon P_X(x), \forall\, x\in\mathcal{X}\Big\}.$$

This notion of typicality, proposed by Orlitsky and Roche [135], is also commonly known as *robust typicality* and is convenient for coding problems with cost constraints or rate-distortion problems. However, it suffers from the deficiency that it is amenable only to *finite* alphabets. This is mitigated by the availability of the *ϵ-weakly typical set* with respect to a distribution $P_X \in \mathcal{P}(\mathcal{X})$, which is defined as

$$\mathcal{A}_\epsilon^{(n)}(P_X) := \left\{x^n \in \mathcal{X}^n : \left|\frac{1}{n}\log\frac{1}{P_X^n(x^n)} - H(P_X)\right| < \epsilon\right\}.$$

When X is a continuous random variable, $H(P_X)$ is to be replaced by the *differential entropy* of X [42]. The conditional versions of these sets can be defined in a natural manner, e.g., the *conditionally ϵ-strongly typical set* of Y given a sequence $x^n \in \mathcal{X}^n$ is

$$\mathcal{T}_\epsilon^{(n)}(P_{XY}|x^n) := \Big\{y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(P_{XY})\Big\}.$$

1.3.5 Asymptotic Notations

Asymptotic notation is used in the usual way [39]. Given two real-valued sequences $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$ and $\{b_n\}_{n=1}^{\infty} \subset \mathbb{R}$, we say that $a_n = O(b_n)$ if $\limsup_{n\to\infty} |a_n/b_n| < \infty$, $a_n = \Omega(b_n)$ if $\liminf_{n\to\infty} |a_n/b_n| > 0$, and $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$. Similarly, $a_n = o(b_n)$ if $\lim_{n\to\infty} |a_n/b_n| = 0$. Finally, if $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are positive sequences, we write $a_n \doteq b_n$ if these sequences are *equal to first-order in the exponent* [42], i.e., $\lim_{n\to\infty} n^{-1}\log(a_n/b_n) = 0$.

1.3.6 Miscellaneous

For two integers m and n, we write $[m:n] = \{m, m+1, \ldots, n\}$ to denote the discrete interval. When $m = 1$, this is abbreviated as $[n]$. Often, for an $R > 0$, we write $[2^{nR}]$ to refer to the set $\{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$. Given a number $a \in [0,1]$, we write $\overline{a} := 1 - a$. Given two numbers $a, b \in [0,1]$, we write $a * b := a\bar{b} + b\bar{a}$ to denote their binary convolution. We write $[a]^+$ to mean $\max\{a, 0\}$ for $a \in \mathbb{R}$. For two bits $a, b \in \{0,1\}$, $a \oplus b$ denotes the binary addition (modulo-2 sum) operation, i.e., $a \oplus b = 0$ if $a = b$ and 1 otherwise. Logarithms are always to the base 2 unless otherwise specified. When we write ln, we are referring to the natural logarithm (to base e).

Vectors are interchangeably denoted by boldface lower case font (e.g., $\mathbf{u}$) or, as mentioned in Section 1.3.1, with a lower case letter and with a superscript indicating its length (e.g., $u^n = (u_1, u_2, \ldots, u_n)$). Matrices (e.g., $\mathbf{M}$) are denoted in boldface upper case font. The i^{th} element of a vector $\mathbf{u}$ is denoted interchangeably as u_i or $[\mathbf{u}]_i$. Similarly, the $(i,j)^{\text{th}}$ element of a matrix $\mathbf{M}$ is denoted interchangeably as $M_{i,j}$ or $[\mathbf{M}]_{i,j}$.

1.4 Mathematical Tools

1.4.1 The Method of Types

We summarize a few key property of types which will turn out to be useful in proving both achievability and converse parts of various common information problems, particularly those with finite alphabets.

For an extensive discussion, the reader is referred to the book by Csiszár and Körner [45].

First, the number of types $|\mathcal{P}_n(\mathcal{X})| \le (n+1)^{|\mathcal{X}|}$ is polynomial in n. Second, for a given type $T \in \mathcal{P}_n(\mathcal{X})$, the size of the type class $(n+1)^{-|\mathcal{X}|} 2^{nH(T)} \le |\mathcal{T}_T| \le 2^{nH(T)}$ is related to the entropy of the type $H(T)$. Third, the Q^n-probability of a sequence $x^n \in \mathcal{T}_T$ is $Q^n(x^n) = 2^{-n(D(T\|Q)+H(T))}$. Consequently, the Q^n-probability of the type class $\mathcal{T}_T$ is bounded as $(n+1)^{-|\mathcal{X}|} 2^{-nD(T\|Q)} \le Q^n(\mathcal{T}_T) \le 2^{-nD(T\|Q)}$.

A particularly useful result that we use repeatedly in Part III of the monograph is *Sanov's theorem* [42], [49], [150], so we review it here.

Theorem 1.1 (Sanov's theorem). Let the components of the random vector $X^n = (X_1, X_2, \ldots, X_n)$ be generated in an independently and identically distributed (i.i.d.) manner from a PMF $Q \in \mathcal{P}(\mathcal{X})$. For any $n \in \mathbb{N}$ and any set of distributions $\mathcal{S} \subset \mathcal{P}(\mathcal{X})$,

$$Q^n(\{x^n : T_{x^n} \in \mathcal{S}\}) \le (n+1)^{|\mathcal{X}|} 2^{-nD(P^*\|Q)},$$

where the *information projection* of Q onto $\mathcal{S}$ is any distribution P^* that satisfies

$$D(P^*\|Q) = \inf_{P \in \mathcal{S}} D(P\|Q).$$

If additionally, $\mathcal{S}$ is equal to the closure of its interior (under the relative topology),[2]

$$\liminf_{n\to\infty} -\frac{1}{n} \log Q^n(\{x^n : T_{x^n} \in \mathcal{S}\}) \ge D(P^*\|Q),$$

and hence,

$$Q^n(\{x^n : T_{x^n} \in \mathcal{S}\}) \doteq 2^{-nD(P^*\|Q)}.$$

Sanov's theorem basically says that the exponent of the probability that the type T_{X^n} of a random sequence $X^n \sim Q^n$ belongs to a set $\mathcal{S}$ is dominated by the relative entropy between the information projection of Q onto $\mathcal{S}$ and Q.

[2] This regularity condition will always be satisfied in the sections to follow.

1.4.2 Couplings

In this monograph, we will often encounter the optimization problems over joint distributions for which their marginals are fixed. Such a joint distribution is known as a coupling. More precisely, a *coupling* P_{XY} of two distributions $Q_X \in \mathcal{P}(\mathcal{Y})$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$ is a joint distribution on $\mathcal{X} \times \mathcal{Y}$ whose X- and Y-marginals are respectively Q_X and Q_Y. The set of all couplings with marginals Q_X and Q_Y is denoted as

$$\mathcal{C}(Q_X, Q_Y) := \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : P_X = Q_X, P_Y = Q_Y\}.$$

Similarly, a conditional coupling $P_{XY|W}$ is a joint conditional distribution whose X- and Y-marginals agree with given marginals $Q_{X|W}$ and $Q_{Y|W}$ respectively. The set of all conditional couplings with marginals $Q_{X|W}$ and $Q_{Y|W}$ is

$$\begin{aligned}&\mathcal{C}(Q_{X|W}, Q_{Y|W})\\ &\quad := \{P_{XY|W} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}|\mathcal{W}) : P_{X|W} = Q_{X|W}, P_{Y|W} = Q_{Y|W}\}.\end{aligned}$$

Couplings have many beautiful properties, but we will not elaborate on them in this monograph; see Thorisson [163] or Yu and Tan [199] for example. One property that is quite remarkable is the *maximal coupling equality* which says that given two distributions Q_X and Q_Y, the total variation distance between them is equal to the probability that X is not equal to Y minimized over all couplings induced by Q_X and Q_Y, i.e.,

$$\min_{P_{XY} \in \mathcal{C}(Q_X, Q_Y)} \Pr(X \neq Y) = |Q_X - Q_Y|.$$

A generalization of the maximal coupling equality that turns out to be useful in the GKW common information problem (Section 3) is stated as follows. This lemma is due to the present authors [199].

Lemma 1.2 (Maximal guessing coupling equality)**.** Given two distributions Q_X and Q_Y, we have

$$\min_{P_{XY} \in \mathcal{C}(Q_X, Q_Y)} \min_{f:\mathcal{X}\to\mathcal{Y}} \Pr\left(Y \neq f(X)\right) = \min_{f:\mathcal{X}\to\mathcal{Y}} \left|Q_Y - Q_{f(X)}\right|. \tag{1.11}$$

The minimization problem on the left-hand side of (1.11) is termed the *maximal guessing coupling problem* (because we would like to maximize the probability that Y is guessed correctly by f acting on X).

The minimization problem on the right-hand side is a classical problem in information theory which is termed the *distribution approximation* or *random number generation* problem [71, Chapter 2]. Lemma 1.2 implies that the maximal guessing coupling problem is equivalent to the distribution approximation problem.

The concept of coupling is naturally involved when we study a problem involving Markov chains, e.g., Wyner's common information and its extensions. One key step to analyze such problems is to simplify multi-letter expressions that involve optimizations over couplings to single-letter ones. This is conveniently facilitated by the *chain rule* on couplings. Before stating this, we first define the *product coupling set*

$$\prod_{i=1}^{n} \mathcal{C}(Q_{X_i|X^{i-1}W}, P_{Y_i|Y^{i-1}W}) := \Bigg\{ \prod_{i=1}^{n} P_{X_iY_i|X^{i-1}Y^{i-1}W} : \\ P_{X_iY_i|X^{i-1}Y^{i-1}W} \in \mathcal{C}(Q_{X_i|X^{i-1}W}, Q_{Y_i|Y^{i-1}W}), i \in [n] \Bigg\}.$$

Lemma 1.3 (Chain Rule for Coupling Sets)**.** For any pair of conditional distributions $(Q_{X^n|W}, Q_{Y^n|W})$, we have

$$\prod_{i=1}^{n} \mathcal{C}(Q_{X_i|X^{i-1}W}, Q_{Y_i|Y^{i-1}W}) \subset \mathcal{C}(Q_{X^n|W}, Q_{Y^n|W}).$$

This lemma can be interpreted as follows. By the usual chain rule for joint distributions, the conditional distributions $Q_{X^n|W}$ and $Q_{Y^n|W}$ can be factorized as $\prod_{i=1}^{n} Q_{X_i|X^{i-1}W}$ and $\prod_{i=1}^{n} Q_{Y_i|Y^{i-1}W}$ respectively. Let $P_{X_iY_i|X^{i-1}Y^{i-1}W}$ be a coupling of each pair of component conditional distributions $(Q_{X_i|X^{i-1}W}, Q_{Y_i|Y^{i-1}W})$. Then, this lemma says that the product of $P_{X_iY_i|X^{i-1}Y^{i-1}W}$ forms a coupling of the product of $Q_{X_i|X^{i-1}W}$ and the product of $Q_{Y_i|Y^{i-1}W}$.

The proof of this lemma can be found in Yu and Tan [204].

Part III

Extensions of Gács–Körner–Witsenhausen's Common Information

8

Non-Interactive Correlation Distillation

In this section, we consider an extension of GKW's common information, termed *Non-Interactive Correlation Distillation*. We recall that GKW's common information measures the amount of "almost identical" randomnesses that can be extracted individually from a pair of correlated sources. By Gács and Körner's theorem [60] (also recall Proposition 3.3), the GKW's common information of a joint source (X, Y) is positive if and only if there exists a pair of non-constant functions (f, g) such that $f(X) = g(Y)$ almost surely. Unfortunately, GKW's common information is zero for many common pairs of sources, such as jointly Gaussian sources and doubly symmetric binary sources (DSBS) with correlation coefficients $\rho \in (-1, 1)$. For these joint sources, even if we wish to extract *a single pair* of identical bits from these sources individually, this innocuous task still turns out to be infeasible.

This observation begs the following natural question: *How can we refine the quantification of common information for these and other sources such that it resembles the GKW's common information and yet is non-zero?* Even though any randomnesses extracted from these sources individually cannot agree almost surely, the extracted randomnesses can indeed agree with a certain probability, which, in this section, we quantify via various probability limit theorems such as the central limit and large deviations theorems. In other words, the extracted randomnesses can be correlated. It is thus natural to quantify the "common information" by the maximal correlation of a pair of random bits that can be extracted from the sources individually. In the literature, determining this maximal correlation is coined the *Noise Stability Problem* (two-set version), the *Non-Interactive Correlation Distillation* or NICD problem. Other names include the *Non-Interactive Binary Simulation Problem* and the *Binary Decision Problem*. This problem was studied by Kamath and Anantharam [94], Yang [188], Mossel *et al.* [124] and Witsenhausen [178] among others.

In this section, we focus mainly on the doubly symmetric binary source (DSBS) parametrized by its correlation coefficient $\rho \in (-1, 1)$. Even though this source is simple, the NICD problem for this source is nontrivial and insights can be drawn from it. In Section 8.1, we define the 2-user NICD problem for the DSBS. Based on the means of the extracted random bits, we define several asymptotic regimes of interest, including the central limit, moderate, and large deviations regimes. In Section 8.2, we discuss various achievability schemes for the NICD problem based on certain geometric structures in Hamming space; these include subcubes and Hamming spheres. These geometric structures are useful to prove existence results in the above-mentioned asymptotic regimes. In Sections 8.3, 8.4, and 8.5 we discuss the optimality of these schemes. Finally, in Section 8.6 we discuss known results in the NICD problem for other sources such as bivariate Gaussians.

8.1 Non-Interactive Correlation Distillation with 2 Users

Consider a doubly symmetric binary distribution π_{XY} on the alphabet $\mathcal{X} \times \mathcal{Y} = \{0, 1\}^2$ with correlation coefficient $\rho \in (0, 1)$, i.e.,

$$\pi_{XY}(x, y) = \begin{cases} \dfrac{1+\rho}{4} & x = y \\ \dfrac{1-\rho}{4} & x \neq y \end{cases}. \tag{8.1}$$

With this parametrization, the *correlation coefficient* of (X, Y), defined in (1.1), is indeed ρ. The pair of random variables $(X, Y) \sim \pi_{XY}$

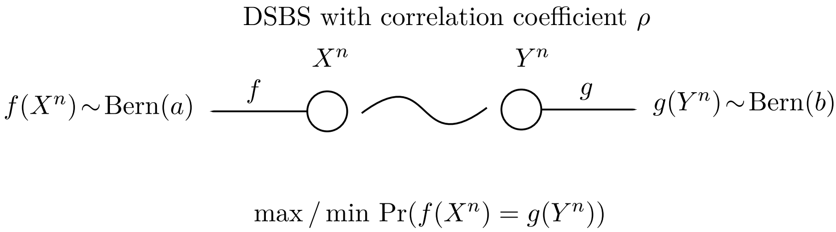

Figure 8.1: The Non-Interactive Correlation Distillation problem with 2 users

corresponds to the DSBS as described in Section 2.3 with *crossover probability* $p = (1-\rho)/2 \in (0, 1/2)$. In this section, we find it convenient to parametrize the DSBS by its correlation coefficient ρ instead of its crossover probability p. It suffices to consider positive ρ as the results carry over to the case for negative ρ by replacing X with $1 - X$. Throughout this section except for Section 8.6, we let (X^n, Y^n) be distributed as the n-fold product distribution π_{XY}^n.

We now introduce the NICD problem with 2 users. This problem is illustrated in Fig. 8.1, in which a source sequence (X^n, Y^n) generated by a DSBS is given, and two random bits $f(X^n)$ and $g(Y^n)$ are generated in a distributed manner using a pair of Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$. The objective of the NICD problem is to maximize or minimize the *agreement probability* of $f(X^n)$ and $g(Y^n)$, i.e., $\Pr(f(X^n) = g(Y^n))$, under the condition that the means of $f(Y^n)$ and $g(Y^n)$ are bounded.

Definition 8.1. Given $a, b \in [0, 1]$, the *forward joint probability* is

$$\overline{\Gamma}^{(n)}(a, b) := \max_{\substack{f,g:\{0,1\}^n \to \{0,1\}: \Pr(f(X^n)=1) \le a, \\ \Pr(g(Y^n)=1) \le b}} \Pr\left(f(X^n) = g(Y^n) = 1\right). \quad (8.2)$$

Similarly, define the *reverse joint probability* as

$$\underline{\Gamma}^{(n)}(a, b) := \min_{\substack{f,g:\{0,1\}^n \to \{0,1\}: \Pr(f(X^n)=1) \ge a, \\ \Pr(g(Y^n)=1) \ge b}} \Pr\left(f(X^n) = g(Y^n) = 1\right). \quad (8.3)$$

In Definition 8.1, we maximize or minimize the probability that both generated bits are equal to one, i.e., $\Pr(f(X^n) = g(Y^n) = 1)$, rather than $\Pr(f(X^n) = g(Y^n))$, since by noting that the marginal probabilities $\Pr(f(X^n) = 1)$ and $\Pr(g(Y^n) = 1)$ are constrained in (8.2) and (8.3), determining the former is equivalent to that of the latter.

8.1.1 Optimizing over Supports of Boolean Functions

Instead of optimizing over the Boolean functions f and g, in the following, we find it convenient for the sake of exploiting the properties of geometric structures (such as Hamming balls and spheres) to optimize over their *supports*. The *support* of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined as the set $\mathcal{A} := \{x^n \in \{0,1\}^n : f(x^n) = 1\}$.

If we denote the supports of f and g as $\mathcal{A}$ and $\mathcal{B}$ respectively, then one can rewrite (8.2) and (8.3) respectively as

$$\overline{\Gamma}^{(n)}(a,b) = \max_{\mathcal{A},\mathcal{B}\subset\{0,1\}^n : \pi_X^n(\mathcal{A})\le a,\, \pi_Y^n(\mathcal{B})\le b} \pi_{XY}^n(\mathcal{A}\times\mathcal{B}), \qquad (8.4)$$

and

$$\underline{\Gamma}^{(n)}(a,b) = \min_{\mathcal{A},\mathcal{B}\subset\{0,1\}^n : \pi_X^n(\mathcal{A})\ge a,\, \pi_Y^n(\mathcal{B})\ge b} \pi_{XY}^n(\mathcal{A}\times\mathcal{B}). \qquad (8.5)$$

Let $\overline{\Gamma}^{(\infty)}$ and $\underline{\Gamma}^{(\infty)}$ respectively denote the pointwise limits of $\overline{\Gamma}^{(n)}$ and $\underline{\Gamma}^{(n)}$ as $n \to \infty$, i.e.,

$$\overline{\Gamma}^{(\infty)}(a,b) := \lim_{n\to\infty} \overline{\Gamma}^{(n)}(a,b) \quad \text{and} \quad \underline{\Gamma}^{(\infty)}(a,b) := \lim_{n\to\infty} \underline{\Gamma}^{(n)}(a,b). \quad (8.6)$$

These are respectively known as the *asymptotic forward* and *asymptotic reverse joint probabilities*.

By definition, the forward and reverse joint probabilities are nondecreasing in each of the parameters when the other is fixed. This implies that there exists an optimal pair of sets $\mathcal{A},\mathcal{B} \subset \{0,1\}^n$ (or Boolean functions (f,g)) attaining the forward joint probability such that

$$\pi_X^n(\mathcal{A}) = \frac{\lfloor a\cdot 2^n\rfloor}{2^n} \quad \text{and} \quad \pi_Y^n(\mathcal{B}) = \frac{\lfloor b\cdot 2^n\rfloor}{2^n}.$$

Indeed, if either of these statements were not true, we can enlarge $\mathcal{A}$ (resp. $\mathcal{B}$) to make its π_X^n-probability (resp. π_Y^n-probability) closer to a (resp. b). Similarly, there exists an optimal pair $(\mathcal{A},\mathcal{B})$ (or Boolean functions (f,g)) attaining the reverse joint probability such that

$$\pi_X^n(\mathcal{A}) = \frac{\lceil a\cdot 2^n\rceil}{2^n} \quad \text{and} \quad \pi_Y^n(\mathcal{B}) = \frac{\lceil b\cdot 2^n\rceil}{2^n}.$$

As a consequence, for dyadic rationals a and b (i.e., $a = M/2^n, b = N/2^n$ with integers $M,N \in \{0,1,\ldots,2^n\}$), the inequalities in the constraints in the definitions of forward and reverse probabilities (i.e., $\overline{\Gamma}^{(n)}(a,b)$ and $\underline{\Gamma}^{(n)}(a,b)$) can be replaced by equalities, without affecting their values. These observations also allow us to conclude that

$$\overline{\Gamma}^{(n)}(1-a,b) = b - \underline{\Gamma}^{(n)}(a,b) \quad \text{for all dyadic rationals } a, b.$$

When we consider the asymptotic case in which $n \to \infty$, i.e., the quantities in (8.6), the requirement that a and b are dyadic rationals can be removed. This implies that for any $a, b \in [0,1]$,

$$\overline{\Gamma}^{(\infty)}(1-a,b) = b - \underline{\Gamma}^{(\infty)}(a,b). \tag{8.7}$$

Hence, for all $(a,b) \in [0,1]^2$, determining the asymptotic forward joint probability in (8.6) is equivalent to determining the asymptotic reverse joint probability and vice versa.

8.1.2 Asymptotic Regimes and Exponents of Interest

The identification of the optimal pairs $(\mathcal{A}, \mathcal{B})$ that attain the forward or reverse joint probabilities in (8.4) and (8.5) constitutes a combinatorial problem and is thus difficult in general. Hence, we focus on the limiting cases as $n \to \infty$ as this simplifies the problem, and the resultant problems are also information-theoretic in nature. Specifically, the following three asymptotic regimes will be considered.

1. Central limit (CL) regime: We set a and b to be constants. We write $a = 2^{-\alpha}$ and $b = 2^{-\beta}$ for a pair of constants $(\alpha, \beta) \in [0, \infty)^2$.

2. Large deviations (LD) regime: We set a and b to be sequences that vanish exponentially fast as $n \to \infty$. In particular, we write $a = 2^{-n\alpha}$ and $b = 2^{-n\beta}$ for a pair of constants $(\alpha, \beta) \in [0,1]^2$.

3. Moderate deviations (MD) regime: We set a and b to be sequences that vanish subexponentially fast as $n \to \infty$. More precisely, $a = 2^{-\theta_n \alpha}, b = 2^{-\theta_n \beta}$ for a pair of constants $(\alpha, \beta) \in [0, \infty)^2$, where $\{\theta_n\}_{n \in \mathbb{N}}$ is a positive sequence satisfying $\theta_n \to \infty$ and $\theta_n / n \to 0$, henceforth called an *MD sequence*.

The MD regime straddles between the CL and LD regimes. It is usually the case if one solves a certain information-theoretic problem in the CL or the LD regimes, a result for the MD regime can be derived as a corollary, for example, by appealing to Taylor's theorem; see Altuğ and Wagner [3], Polyanskiy and Verdú [140], and Tan [159] for example. We will see that this is also the case for the NICD problem.

In the following section, we will set $\mathcal{A}$ and $\mathcal{B}$ to be subcubes, Hamming balls, and Hamming spheres. These are prototypical subsets in the Hamming space that are amenable to analyses. We will then apply various probabilistic limit theorems—such as the central limit theorem and large and moderate deviations theorems—to derive the "performances" of these subsets in attaining the forward and reverse joint probabilities. We formally define several exponents of interest.

Definition 8.2. Consider the following exponents:

1. Forward and reverse CL exponents: For $\alpha, \beta \in [0, \infty)$,

$$\underline{\Upsilon}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \overline{\Gamma}^{(n)}(2^{-\alpha}, 2^{-\beta}) \quad \text{and} \tag{8.8}$$
$$\overline{\Upsilon}_{\mathrm{CL}}^{(n)}(\alpha, \beta) := -\log \underline{\Gamma}^{(n)}(2^{-\alpha}, 2^{-\beta}).$$

2. Forward and reverse LD exponents: For $\alpha, \beta \in [0, 1]$,

$$\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \overline{\Gamma}^{(n)}(2^{-n\alpha}, 2^{-n\beta}) \quad \text{and} \tag{8.9}$$
$$\overline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \underline{\Gamma}^{(n)}(2^{-n\alpha}, 2^{-n\beta}). \tag{8.10}$$

3. Forward and reverse MD exponents: Given an MD sequence $\{\theta_n\}$, and for $\alpha, \beta \in [0, \infty)$,

$$\underline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha, \beta) := -\frac{1}{\theta_n} \log \overline{\Gamma}^{(n)}(2^{-\theta_n \alpha}, 2^{-\theta_n \beta}) \quad \text{and} \tag{8.11}$$
$$\overline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha, \beta) := -\frac{1}{\theta_n} \log \underline{\Gamma}^{(n)}(2^{-\theta_n \alpha}, 2^{-\theta_n \beta}). \tag{8.12}$$

4. Define $\underline{\Upsilon}_{\mathrm{CL}}^{(\infty)}$, $\overline{\Upsilon}_{\mathrm{CL}}^{(\infty)}$, $\underline{\Upsilon}_{\mathrm{LD}}^{(\infty)}$, $\overline{\Upsilon}_{\mathrm{LD}}^{(\infty)}$, $\underline{\Upsilon}_{\mathrm{MD}}^{(\infty)}$, and $\overline{\Upsilon}_{\mathrm{MD}}^{(\infty)}$ as the pointwise limits of the above exponents as $n \to \infty$.

The reader may notice that the definitions in (8.8)–(8.12) appear to be redundant, since each of the forward (resp. reverse) exponents is equivalent to the forward (resp. reverse) joint probability in the sense that if the forward (resp. reverse) joint probability has been determined, then each of the forward (resp. reverse) exponents has also been determined. This also means the forward (resp. reverse) exponents are also "equivalent". For example, for each $n \in \mathbb{N}$, $\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha,\beta) = \frac{1}{n}\underline{\Upsilon}_{\mathrm{CL}}^{(n)}(n\alpha, n\beta)$ and $\underline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha,\beta) = \frac{1}{\theta_n}\underline{\Upsilon}_{\mathrm{CL}}^{(n)}(\theta_n\alpha, \theta_n\beta)$. We introduce these notations because in the sequel, we will introduce several *dimension-free* bounds (e.g., Theorem 8.9) that can be conveniently expressed in terms of the exponents defined in (8.8)–(8.12). Here, a dimension-free bound is one that is independent of the dimension (or blocklength) n, but is valid for all dimensions n.

In the following, we introduce bounds on the NICD exponents in (8.8)–(8.12). As is conventional in information theory, there are two parts to this endeavor. In the achievability part that will be discussed in Section 8.2, we construct subsets $\mathcal{A}$ and $\mathcal{B}$ that upper bound the forward exponents and lower bound the reverse exponents. In the converse parts that will be discussed in Section 8.3–8.5, we demonstrate impossibility results, i.e., lower bounds on the forward exponents and upper bounds on the reverse exponents. The achievability and converse bounds match in some special cases.

8.2 Achievability: Subcubes, Hamming Balls, and Spheres

We now consider the achievability parts, i.e., deriving lower bounds for the forward joint probability and upper bounds for the reverse joint probability. For these parts, we consider three canonical types of subsets in Hamming space—subcubes, Hamming balls, and Hamming spheres.

8.2.1 Subcubes

An $(n-k)$-*subcube* $\mathbb{C}_{n-k}$ is a set of vectors $x^n \in \{0,1\}^n$ with k components held fixed. For example, if we fix the first k components to 1, then we get the $(n-k)$-subcube $\{1^k\} \times \{0,1\}^{n-k}$, where 1^k denotes the length-k all-ones vector. For any set $\mathcal{A} \subset \{0,1\}^n$, we say that its *indicator*, denoted as $\mathbb{1}_{\mathcal{A}}$, is the function $f : \{0,1\}^n \to \{0,1\}$ such that $f(x^n) = 1$ for all $x^n \in \mathcal{A}$ and $f(x^n) = 0$ for all $x^n \notin \mathcal{A}$. The indicator of

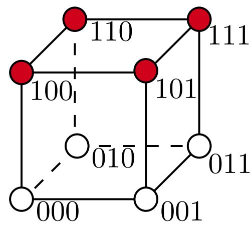

Figure 8.2: A subcube (shaded) in $\{0,1\}^3$ with the first component fixed to 1

the subcube $\{1^k\} \times \{0,1\}^{n-k}$ is $x^n \in \{0,1\}^n \mapsto \prod_{i=1}^k x_i$. An important class of subcubes is the class of $(n-1)$-subcubes, e.g., $\{1\} \times \{0,1\}^{n-1}$. An $(n-1)$-subcube with $n = 3$ is illustrated in Fig. 8.2. The indicators of $(n-1)$-subcubes are the functions $x^n \mapsto x_i$ or $x^n \mapsto 1 - x_i$ for $i \in [n]$. Such functions are known as *dictator* functions.

We now return to the NICD problem. For $a = b = 2^{-k}$ for a positive integer k, we choose $\mathcal{A}$ and $\mathcal{B}$ as a pair of identical $(n-k)$-subcubes. By referring to the joint distribution in (8.1), we see that the joint probability induced by $(\mathcal{A}, \mathcal{B})$ is

$$\pi_{XY}^n(\mathcal{A} \times \mathcal{B}) = \pi_{XY}(1,1)^k = \Big(\frac{1+\rho}{4}\Big)^k. \tag{8.13}$$

On the other hand, if we choose $\mathcal{A}$ and $\mathcal{B}$ as a pair of anti-symmetric $(n-k)$-subcubes, i.e., $\mathcal{A} = 1^n - \mathcal{B} = \mathcal{C}_{n-k}$, then the induced joint probability is

$$\pi_{XY}^n(\mathcal{A} \times \mathcal{B}) = \pi_{XY}(1,0)^k = \Big(\frac{1-\rho}{4}\Big)^k. \tag{8.14}$$

For the more general case in which $a = 2^{-k_1}$ and $b = 2^{-k_2}$ for integers $0 \le k_1 \le k_2$, if we choose $(\mathcal{A}, \mathcal{B})$ as a pair of "nested" subcubes, i.e., $\mathcal{A} = \{1^{k_1}\} \times \{0,1\}^{n-k_1}$ and $\mathcal{B} = \{1^{k_2}\} \times \{0,1\}^{n-k_2}$, then the induced joint probability

$$\pi_{XY}^n(\mathcal{A} \times \mathcal{B}) = \Big(\frac{1}{2}\Big)^{k_2-k_1}\Big(\frac{1+\rho}{4}\Big)^{k_1}.$$

For the same case, if we choose $(\mathcal{A}, \mathcal{B})$ as a pair of "anti-nested" subcubes, i.e., $\mathcal{A} = \{1^{k_1}\} \times \{0,1\}^{n-k_1}$ and $\mathcal{B} = \{0^{k_2}\} \times \{0,1\}^{n-k_2}$, then

$$\pi_{XY}^n(\mathcal{A} \times \mathcal{B}) = \Big(\frac{1}{2}\Big)^{k_2-k_1}\Big(\frac{1-\rho}{4}\Big)^{k_1}.$$

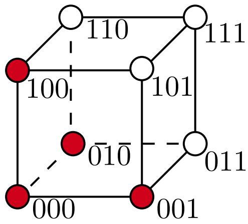

Figure 8.3: A Hamming ball (shaded) in $\{0,1\}^3$ centered at $(0,0,0)$ with radius 1

We now discuss the case in which a and b are dyadic rationals (i.e., $a = M/2^n, b = N/2^n$ for some integers M, N). Observe that if a dyadic rational a is not equal to 2^{-k} for some integer k, then there is no subcube with π_X^n-probability *exactly* equal to a. Hence, to achieve better performances, a generalization of subcubes $\{0^k\} \times \{0,1\}^{n-k}$ and $\{1^k\} \times \{0,1\}^{n-k}$, called *lexicographic sets*, turns out to be useful. A subset of $\{0,1\}^n$ is called *lexicographic* if the elements are selected as the first sequences in some lexicographic order (either ascending or descending). A Boolean function is called *lexicographic* if its support is a lexicographic set. By setting $\mathcal{A}$ and $\mathcal{B}$ to be two lexicographic sets both in ascending (or descending) order, we can obtain a relatively large joint probability $\pi_{XY}^n(\mathcal{A} \times \mathcal{B})$. On the other hand, if we set $\mathcal{A}$ and $\mathcal{B}$ to be two lexicographic sets such that one is chosen in ascending order and the other in descending order, we can obtain a relatively small joint probability $\pi_{XY}^n(\mathcal{A} \times \mathcal{B})$. The explicit expressions for these two joint probabilities are complicated, and thus we omit them. A lexicographic set chosen in ascending order can then be written as $\{x^n \in \{0,1\}^n : \sum_{i=1}^n 2^{i-1} x_i \le r\}$ for some r. This is a special case of so-called *linear threshold functions*, which is discussed in detail in [131].

8.2.2 Hamming Balls

A *Hamming ball* centered at $y^n \in \{0,1\}^n$ with radius $r \in \{0,1,\ldots,n\}$ takes the form $\mathbb{B}_r(y^n) := \{x^n \in \{0,1\}^n : d_{\mathrm{H}}(x^n, y^n) \le r\}$, where $d_{\mathrm{H}}(x^n, y^n) := \sum_{i=1}^n \mathbb{1}\{x_i \ne y_i\}$ denotes the *Hamming distance* between vectors x^n and y^n. An example of a Hamming ball with radius 1 is illustrated in Fig. 8.3. In the following, we only consider Hamming balls that are centered at $0^n = (0,0,\ldots,0)$ or $1^n = (1,1,\ldots,1)$. For these

Hamming balls (with radius r), we can rewrite them as $\{x^n \in \{0,1\}^n : \sum_{i=1}^n x_i \le r\}$ and $\{x^n \in \{0,1\}^n : \sum_{i=1}^n x_i \ge n-r\}$ respectively.

We now set $\mathcal{A}$ and $\mathcal{B}$ in the NICD problem to be Hamming balls. We first consider the CL regime in which we choose $\mathcal{A}$ and $\mathcal{B}$ to be a pair of *concentric* Hamming balls. More specifically, $\mathcal{A}_n := \mathbb{B}_{r_n}(0^n)$ and $\mathcal{B}_n = \mathbb{B}_{s_n}(0^n)$ for some sequences $\{r_n\}_{n\in\mathbb{N}}$ and $\{s_n\}_{n\in\mathbb{N}}$. We append the subscript n to $\mathcal{A}$ and $\mathcal{B}$, to indicate that these two sets depend on n. We can rewrite $\mathcal{A}_n$ as $\{x^n : \sum_{i=1}^n x_i \le r_n\}$. Hence, the marginal probability $\pi_X^n(\mathcal{A}_n)$ can be written as $\Pr(\sum_{i=1}^n X_i \le r_n)$ where $\{X_i\}_{i=1}^n$ are i.i.d. with each $X_i \sim \mathrm{Bern}(1/2)$. To calculate the limiting value of this probability as $n \to \infty$, one may apply several well-known concentration of measure theorems, including the central limit theorem or various large deviations theorems. Since we focus on the CL regime here, we require that $\pi_X^n(\mathcal{A}_n)$ tends to a non-vanishing constant. Hence, we set the radius $r_n = \frac{n}{2} + \frac{\lambda\sqrt{n}}{2}$ for some $\lambda \in \mathbb{R}$. Then, the (univariate) central limit theorem yields

$$\lim_{n\to\infty} \pi_X^n(\mathcal{A}_n) = \Phi(\lambda), \tag{8.15}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard univariate Gaussian distribution. Similarly, if we set the radius $s_n = \frac{n}{2} + \frac{\mu\sqrt{n}}{2}$ for some $\mu \in \mathbb{R}$, we obtain

$$\lim_{n\to\infty} \pi_Y^n(\mathcal{B}_n) = \Phi(\mu).$$

We now estimate the asymptotic value of the joint probability $\pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n)$ where $\mathcal{A}_n$ and $\mathcal{B}_n$ are concentric spheres with radii r_n and s_n respectively. Note that this probability can be restated as $\Pr(\sum_{i=1}^n X_i \le r_n, \sum_{i=1}^n Y_i \le s_n)$ where $(X^n, Y^n) = \{(X_i, Y_i)\}_{i=1}^n$ is a source sequence generated by a DSBS with correlation coefficient ρ. The multivariate central limit theorem then yields

$$\lim_{n\to\infty} \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) = \Phi_\rho(\lambda, \mu), \tag{8.16}$$

where $\Phi_\rho(\cdot,\cdot)$ is the joint CDF of the zero-mean bivariate Gaussian distribution with covariance matrix

$$\mathbf{K} := \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{8.17}$$

Based on the asymptotic results in (8.15)–(8.16), one can obtain a lower bound on the forward joint probability in the NICD problem [131, Ex. 9.24 and 10.5].

Proposition 8.1. For $a, b \in (0,1)$,

$$\overline{\Gamma}^{(\infty)}(a,b) \geq \Lambda_\rho(a,b), \tag{8.18}$$

where

$$\Lambda_\rho(a,b) := \Phi_\rho(\Phi^{-1}(a), \Phi^{-1}(b)). \tag{8.19}$$

Here $\Lambda_\rho(\cdot,\cdot)$ is known as the *bivariate normal copula* or the *Gaussian quadrant probability function*. Thanks to the equivalence between the forward and reverse joint probabilities as stated in (8.7), (8.18) can alternatively be expressed in terms of the reverse joint probability as

$$\underline{\Gamma}^{(\infty)}(a,b) \leq \Lambda_{-\rho}(a,b). \tag{8.20}$$

The upper bound $\Lambda_{-\rho}(a,b)$ is achieved by a sequence of pairs of anti-concentric balls $\mathcal{A}_n = \mathbb{B}_{r_n}(0^n)$ and $\mathcal{B}_n = \mathbb{B}_{s_n}(1^n)$.

Considering the exponents of the probabilities in (8.18) and (8.20),

$$\underline{\Upsilon}_{\mathrm{CL}}^{(\infty)}(\alpha,\beta) \leq \underline{\Upsilon}_{\mathrm{CL}}(\alpha,\beta) := -\log \Lambda_\rho(2^{-\alpha}, 2^{-\beta}) \quad \text{and} \tag{8.21}$$

$$\overline{\Upsilon}_{\mathrm{CL}}^{(\infty)}(\alpha,\beta) \geq \overline{\Upsilon}_{\mathrm{CL}}(\alpha,\beta) := -\log \Lambda_{-\rho}(2^{-\alpha}, 2^{-\beta}). \tag{8.22}$$

We next consider the LD and MD regimes. Although it is certainly possible to set $\mathcal{A}_n$ and $\mathcal{B}_n$ to be Hamming balls to obtain achievability results for these two regimes, we prefer not to do so here. This is because, it is much easier to derive the same results by using Hamming *spheres* or *spherical shells*. Therefore, we consider the LD and MD regimes in the following subsection after we introduce Hamming spheres.

8.2.3 Hamming Spheres

A *Hamming sphere* centered at $y^n \in \{0,1\}^n$ with radius $r \in \{0,1,\ldots,n\}$ takes the form $\mathbb{S}_r(y^n) := \{x^n \in \{0,1\}^n : d_{\mathrm{H}}(x^n, y^n) = r\}$. See Fig. 8.4 for an illustration. The definition of Hamming spheres differs from that for Hamming balls in the condition $d_{\mathrm{H}}(x^n, y^n) = r$ in which *equality* is mandated. Similarly to the previous subsection, here we

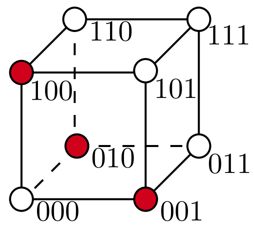

Figure 8.4: A Hamming sphere (shaded) in $\{0,1\}^3$ centered at $(0,0,0)$ with radius 1

also only consider Hamming spheres centered at either 0^n or 1^n, for which we can rewrite them respectively as $\{x^n : \sum_{i=1}^n x_i = r\}$ and $\{x^n : \sum_{i=1}^n x_i = n-r\}$. These Hamming spheres can be regarded as *type classes* with types $(\bar{\lambda}, \lambda)$ and $(\lambda, \bar{\lambda})$ respectively in Hamming space, where $\lambda := \frac{r}{n}$ and $\bar{\lambda} := 1-\lambda$. Observe that $\mathbb{S}_r(0^n)$ is the same as $\mathbb{S}_{n-r}(1^n)$. Notwithstanding this equivalence, we term a pair of spheres $\mathbb{S}_{r_1}(0^n)$ and $\mathbb{S}_{r_2}(0^n)$ as a pair of *concentric* spheres if $r_1, r_2 \le n/2$ or $r_1, r_2 \ge n/2$, and as a pair of *anti-concentric* spheres if $r_1 \le n/2 \le r_2$ or $r_2 \le n/2 \le r_1$.

For the LD regime, we choose $\mathcal{A}_n$ and $\mathcal{B}_n$ to be a pair of concentric or anti-concentric Hamming spheres, i.e., $\mathcal{A}_n = \mathbb{S}_{r_n}(0^n)$ and $\mathcal{B}_n = \mathbb{S}_{s_n}(0^n)$ with $r_n = \lfloor \lambda n \rfloor$ or $\lceil \lambda n \rceil$ and $s_n = \lfloor \mu n \rfloor$ or $\lceil \mu n \rceil$, where $\lambda, \mu \in [0,1]$. By Sanov's theorem [49] (stated in Theorem 1.1),

$$\lim_{n\to\infty} -\frac{1}{n} \log \pi_X^n(\mathcal{A}_n) = D((\bar{\lambda}, \lambda) \| \pi_X) \quad \text{and}$$
$$\lim_{n\to\infty} -\frac{1}{n} \log \pi_Y^n(\mathcal{B}_n) = D((\bar{\mu}, \mu) \| \pi_Y).$$

Since X is uniform on $\{0,1\}$, we can write $D((\bar{\lambda}, \lambda) \| \pi_X) = 1 - h(\lambda)$.

For the joint probability, observe that the set $\mathcal{A}_n \times \mathcal{B}_n$ is a union of joint type classes with types T_{XY} satisfying the condition that its marginals T_X and T_Y are equal to $(\bar{\lambda}, \lambda)$ and $(\bar{\mu}, \mu)$ respectively. Hence, by Sanov's theorem, the joint probability satisfies

$$\lim_{n\to\infty} -\frac{1}{n} \log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) = \mathsf{D}((\bar{\lambda}, \lambda), (\bar{\mu}, \mu) \| \pi_{XY}),$$

where, in analogy to Definition 4.1, the *minimal relative entropy* with respect to π_{XY} over all couplings of Q_X and Q_Y is defined as

$$\mathsf{D}(Q_X, Q_Y \| \pi_{XY}) := \min_{Q_{XY} \in \mathcal{C}(Q_X, Q_Y)} D(Q_{XY} \| \pi_{XY}). \tag{8.23}$$

Optimizing the exponent $\mathsf{D}((\bar{\lambda}, \lambda), (\bar{\mu}, \mu) \| \pi_{XY})$ over all feasible pairs of (λ, μ), yields the following achievability result.

Proposition 8.2. For all $\alpha, \beta \in (0, 1)$,

$$\begin{aligned} \underline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) &\le \underline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) \\ &:= \min_{\substack{Q_X, Q_Y: \\ D(Q_X \| \pi_X) \ge \alpha,\, D(Q_Y \| \pi_Y) \ge \beta}} \mathsf{D}(Q_X, Q_Y \| \pi_{XY}), \end{aligned} \tag{8.24}$$

and

$$\begin{aligned} \overline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) &\ge \overline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) \\ &:= \min_{\substack{Q_X, Q_Y: \\ D(Q_X \| \pi_X) \le \alpha,\, D(Q_Y \| \pi_Y) \le \beta}} \mathsf{D}(Q_X, Q_Y \| \pi_{XY}). \end{aligned} \tag{8.25}$$

The bounds in (8.24) and (8.25) are attained by sequences of concentric and anti-concentric Hamming spheres respectively. By the method of types, it is easy to observe that they also can be respectively attained by sequences of concentric and anti-concentric *balls* (since a Hamming ball consists of several spheres and there is one sphere that dominates the others in the sense of the exponent). The above inequalities were conjectured to be tight by Ordentlich, Polyanskiy, and Shayevitz [133]. We refer to this as the *OPS conjecture* in the sequel.

Conjecture 8.1 (OPS Conjecture)**.** For the DSBS and $\alpha, \beta \in (0, 1)$,

$$\underline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \stackrel{?}{=} \underline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) \quad \text{and} \quad \overline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha, \beta) \stackrel{?}{=} \overline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta).$$

In Section 8.5, we discuss the optimality of Hamming spheres in the LD regime, leading to the proof this conjecture. However, before doing this, we first focus on achievability results by Hamming spherical shells in the MD regime.

For the MD regime, we choose the sets in the NICD problem to be two spherical shells (annuli), with thickness in the order of $\sqrt{n\theta_n}$.

Specifically, for a fixed and small $\epsilon > 0$, we choose

$$\mathcal{A}_n = \bigcup_{r \in n/2 + [\lambda, \lambda+\epsilon]\sqrt{n\theta_n}} \mathbb{S}_r(0^n) \quad \text{and} \quad \mathcal{B}_n = \bigcup_{s \in n/2 + [\mu, \mu+\epsilon]\sqrt{n\theta_n}} \mathbb{S}_s(0^n),$$

where $\{\theta_n\}$ is an MD sequence, and $\lambda, \mu \in \mathbb{R}$. In other words, we choose $\mathcal{A}_n$ and $\mathcal{B}_n$ to be unions of type classes induced by types $Q_X = \pi_X + \sqrt{\theta_n/n}\,\eta_X$ and $Q_Y = \pi_Y + \sqrt{\theta_n/n}\,\eta_Y$ respectively, where η_X and η_Y are functions such that $\sum_{x \in \{0,1\}} \eta_X(x) = 0$ and $\sum_{y \in \{0,1\}} \eta_Y(y) = 0$ and $\eta_X(1) \in [\lambda, \lambda+\epsilon]$ and $\eta_Y(1) \in [\mu, \mu+\epsilon]$. Let

$$\hat{\chi}^2(\eta \| \pi) := \sum_{x \in \{0,1\}} \frac{\eta(x)^2}{\pi(x)}.$$

and notice that $\hat{\chi}^2(Q - \pi \| \pi)$ is the chi-squared divergence from Q to π. In analogy to the minimal relative entropy in (8.23), we define

$$\hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}) := \inf_{\eta_{XY} \in \overline{\mathcal{C}}(\eta_X, \eta_Y)} \hat{\chi}^2(\eta_{XY} \| \pi_{XY}),$$

where $\overline{\mathcal{C}}(\eta_X, \eta_Y)$ is the set of all bivariate functions $\eta_{XY} : \{0,1\}^2 \to \mathbb{R}$ such that their X- and Y-marginals are equal to η_X and η_Y respectively and $\sum_{x,y} \eta_{XY}(x, y) = 0$. Then, letting $\theta_n \to \infty$ and then $\epsilon \downarrow 0$, by the moderate deviations theorem [49], [181],

$$\lim_{n \to \infty} -\frac{1}{\theta_n} \log \pi_X^n(\mathcal{A}_n) = \frac{1}{2} \hat{\chi}^2(\eta_X \| \pi_X), \tag{8.26}$$

$$\lim_{n \to \infty} -\frac{1}{\theta_n} \log \pi_Y^n(\mathcal{B}_n) = \frac{1}{2} \hat{\chi}^2(\eta_Y \| \pi_Y), \quad \text{and}$$

$$\lim_{n \to \infty} -\frac{1}{\theta_n} \log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) = \frac{1}{2} \hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}). \tag{8.27}$$

In fact, (8.27) requires the continuity of $(\eta_X, \eta_Y) \mapsto \hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY})$; this follows from the following lemma.

Lemma 8.1. For $\eta_X = (-\lambda, \lambda)$ and $\eta_Y = (-\mu, \mu)$, we have

$$\hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}) = \frac{2(\lambda+\mu)^2}{1+\rho} + \frac{2(\lambda-\mu)^2}{1-\rho}. \tag{8.28}$$

Proof. One can calculate that the optimal η_{XY} attaining the maximum in the definition of $\hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY})$ is

$$\eta_{XY} = \begin{bmatrix} p - \lambda - \mu & \mu - p \\ \lambda - p & p \end{bmatrix},$$

where $p = (\lambda + \mu)/2$. Hence, (8.28) follows. ∎

Optimizing the exponent $\frac{1}{2}\hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY})$ over all feasible $\eta_X = (-\lambda, \lambda)$ and $\eta_Y = (-\mu, \mu)$ yields the following proposition.

Proposition 8.3. For $\alpha, \beta > 0$,

$$\underline{\Upsilon}_{\mathrm{MD}}^{(\infty)}(\alpha, \beta) \le \underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) := \inf \hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}) \quad \text{and} \tag{8.29}$$

$$\overline{\Upsilon}_{\mathrm{MD}}^{(\infty)}(\alpha, \beta) \ge \overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) := \sup \hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}). \tag{8.30}$$

where the inf in (8.29) is over the set of functions $\eta_X, \eta_Y : \{0, 1\} \to \mathbb{R}$ such that $\sum_x \eta_X(x) = \sum_y \eta_Y(y) = 0$ and

$$\hat{\chi}^2(\eta_X \| \pi_X) \ge \alpha \quad \text{and} \quad \hat{\chi}^2(\eta_Y \| \pi_Y) \ge \beta, \tag{8.31}$$

and the sup in (8.30) is over the same set of functions (η_X, η_Y) but with the directions of the inequalities in (8.31) reversed.

The bounds in (8.29) and (8.30) are respectively attained by sequences of concentric and anti-concentric Hamming spheres or balls. The reader may have noticed that the constant 1/2 in (8.26)–(8.27) has been removed in (8.29) and (8.30). This is because, by definition, $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$ are *homogeneous* (of degree 1), i.e., for any $\gamma > 0$,

$$\underline{\Upsilon}_{\mathrm{MD}}(\gamma\alpha, \gamma\beta) = \gamma \underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) \quad \text{and} \tag{8.32}$$

$$\overline{\Upsilon}_{\mathrm{MD}}(\gamma\alpha, \gamma\beta) = \gamma \overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta). \tag{8.33}$$

The bounds in (8.29) and (8.30) can be further simplified as follows.

Lemma 8.2. For $\alpha, \beta > 0$,

$$\underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \begin{cases} \dfrac{\alpha + \beta - 2\rho\sqrt{\alpha\beta}}{1 - \rho^2} & \rho^2\alpha \le \beta \le \frac{\alpha}{\rho^2} \\ \alpha & \beta < \rho^2\alpha \\ \beta & \alpha < \rho^2\beta \end{cases} \quad \text{and} \tag{8.34}$$

$$\overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \frac{\alpha + \beta + 2\rho\sqrt{\alpha\beta}}{1 - \rho^2}. \tag{8.35}$$

Proof. Observe by the uniformity of π_X and π_Y that $\hat{\chi}^2(\eta_X \| \pi_X) = 4\lambda^2$ and $\hat{\chi}^2(\eta_Y \| \pi_Y) = 4\mu^2$. Combining these with Lemma 8.1 yields that

$$\underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \min_{\lambda,\mu: 4\lambda^2 \ge \alpha, 4\mu^2 \ge \beta} \frac{2(\lambda + \mu)^2}{1 + \rho} + \frac{2(\lambda - \mu)^2}{1 - \rho} \quad \text{and}$$

$$\overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \max_{\lambda,\mu: 4\lambda^2 \le \alpha, 4\mu^2 \le \beta} \frac{2(\lambda + \mu)^2}{1 + \rho} + \frac{2(\lambda - \mu)^2}{1 - \rho}.$$

By the rearrangement inequality and by symmetry, it suffices to consider $\lambda, \mu \geq 0$ for $\underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta)$ and $\lambda \leq 0 \leq \mu$ for $\overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta)$. This results in

$$\underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \min_{\lambda \geq \frac{\sqrt{\alpha}}{2}, \mu \geq \frac{\sqrt{\beta}}{2}} \frac{2(\lambda+\mu)^2}{1+\rho} + \frac{2(\lambda-\mu)^2}{1-\rho} \quad \text{and} \tag{8.36}$$

$$\overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) = \max_{-\frac{\sqrt{\alpha}}{2} \leq \lambda \leq 0 \leq \mu \leq \frac{\sqrt{\beta}}{2}} \frac{2(\lambda+\mu)^2}{1+\rho} + \frac{2(\lambda-\mu)^2}{1-\rho}. \tag{8.37}$$

By calculus, one can verify that the right-hand sides of (8.36) and (8.37) are respectively equal to the right-hand sides of (8.34) and (8.35). □

We conclude this section by discussing the relationships between the MD and CL exponents as well as the MD and LD exponents. We can recover the MD exponents from the CL or LD exponents if the MD sequence $\{\theta_n\}$ additionally satisfies $(\log n)/\theta_n \to 0$ as $n \to \infty$. Roughly speaking, in the MD regime, we chose the radii r_n and s_n of Hamming spheres such that $\frac{r_n}{n} \approx \frac{1}{2} + \lambda\sqrt{\epsilon}$ and $\frac{s_n}{n} \approx \frac{1}{2} + \mu\sqrt{\epsilon}$, where $\epsilon := \frac{\theta_n}{n} \to 0$ as $n \to \infty$. This implies that the types corresponding to the spheres are $Q_X \approx \pi_X + \sqrt{\epsilon}\,\eta_X$ and $Q_Y \approx \pi_Y + \sqrt{\epsilon}\,\eta_Y$ as $\epsilon \downarrow 0$. Note that in Sanov's theorem, the LD exponent of the probability of a Hamming sphere with type Q_X is $D(Q_X \| \pi_X) + O(\frac{\log n}{n})$. Hence, if the MD sequence $\{\theta_n\}$ additionally satisfies $(\log n)/\theta_n \to 0$ as $n \to \infty$, this LD exponent is dominated by the term $D(Q_X \| \pi_X)$, which allows us to omit the $O(\frac{\log n}{n})$ term. Moreover, by Taylor's theorem,

$$D(Q_X \| \pi_X) = \frac{\epsilon}{2}\hat{\chi}^2(\eta_X \| \pi_X) + o(\epsilon),$$
$$D(Q_Y \| \pi_Y) = \frac{\epsilon}{2}\hat{\chi}^2(\eta_Y \| \pi_Y) + o(\epsilon),$$

and similarly,

$$\mathsf{D}(Q_X, Q_Y \| \pi_{XY}) = \frac{\epsilon}{2}\hat{\mathsf{X}}^2(\eta_X, \eta_Y \| \pi_{XY}) + o(\epsilon) \quad \text{as } \epsilon \downarrow 0.$$

We obtain the MD exponents by replacing D and D in the LD exponents with $\frac{\epsilon}{2}\hat{\chi}^2$ and $\frac{\epsilon}{2}\hat{\mathsf{X}}^2$ respectively. Formally,

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \underline{\Upsilon}_{\mathrm{LD}}(\epsilon\alpha, \epsilon\beta) = \underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) \quad \text{and} \tag{8.38}$$

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \overline{\Upsilon}_{\mathrm{LD}}(\epsilon\alpha, \epsilon\beta) = \overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta). \tag{8.39}$$

Furthermore, the MD exponents can be also recovered from the CL exponents. By the Berry–Esseen theorem [19], [54], under the condition that the MD sequence $\{\theta_n\}$ satisfies $(\log n)/\theta_n \to 0$ as $n \to \infty$, the probability of a Hamming ball is dominated by the term involving the Gaussian cumulative distribution function $\Phi(\cdot)$. In other words, the additive error term in the Berry–Esseen theorem, which scales as $O(\frac{1}{\sqrt{n}})$, is negligible asymptotically. On the other hand, O'Donnell [131, Ex. 9.24 and 10.5] shows that

$$\lim_{\theta\to\infty} \frac{1}{\theta} \underline{\Upsilon}_{\mathrm{CL}}(\theta\alpha, \theta\beta) = \underline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta) \quad \text{and} \tag{8.40}$$

$$\lim_{\theta\to\infty} \frac{1}{\theta} \overline{\Upsilon}_{\mathrm{CL}}(\theta\alpha, \theta\beta) = \overline{\Upsilon}_{\mathrm{MD}}(\alpha, \beta), \tag{8.41}$$

where $\underline{\Upsilon}_{\mathrm{CL}}$ and $\overline{\Upsilon}_{\mathrm{CL}}$ are defined in (8.21) and (8.22) respectively.

8.2.4 Numerical Results and Comparisons

We now evaluate the various exponents for the DSBS with correlation coefficient ρ. Define $\kappa := (\frac{1+\rho}{1-\rho})^2$,

$$D_{a,b}(p) := D\left(\begin{bmatrix} 1+p-a-b & b-p \\ a-p & p \end{bmatrix} \middle\| \pi_{XY}\right) \quad \text{and}$$
$$\mathsf{D}(a,b) := \min_{\max\{0,a+b-1\}\le p\le \min\{a,b\}} D_{a,b}(p) = D_{a,b}(p^*_{a,b}),$$

where $h(\cdot)$ is the binary entropy function, and

$$p^*_{a,b} := \frac{(\kappa-1)(a+b)+1-\sqrt{((\kappa-1)(a+b)+1)^2-4\kappa(\kappa-1)ab}}{2(\kappa-1)}.$$

For the DSBS, $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$, defined in (8.24) and (8.25), respectively can be written in closed form as

$$\underline{\Upsilon}_{\mathrm{LD}}(\alpha,\beta) = \mathsf{D}(h^{-1}(1-\alpha), h^{-1}(1-\beta)) \quad \text{and}$$
$$\overline{\Upsilon}_{\mathrm{LD}}(\alpha,\beta) = \mathsf{D}(h^{-1}(1-\alpha), 1-h^{-1}(1-\beta)),$$

where $h^{-1} : [0,1] \to [0,1/2]$ is the inverse of the binary entropy function h when its domain is restricted to $[0,1/2]$.

We plot the CL exponents achieved by Hamming balls, and the MD and LD exponents achieved by Hamming balls, spheres, or spherical

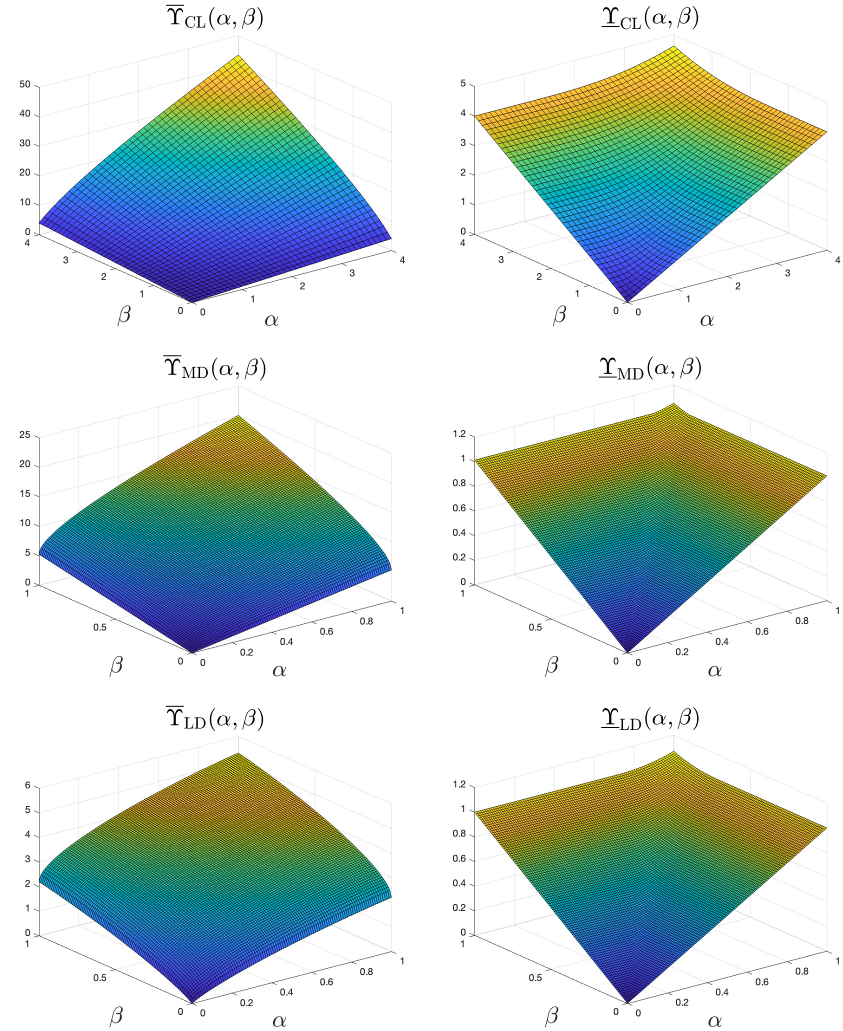

Figure 8.5: Forward and reverse CL, MD, and LD exponents induced by Hamming balls (or spheres) for $\rho = 0.9$. Observe that $\underline{\Upsilon}_{\mathrm{MD}}$ and $\underline{\Upsilon}_{\mathrm{LD}}$ appear to be convex while $\overline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ appear to be concave. The convexity and concavity of $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ respectively have implications for the OPS conjecture (Conjecture 8.1) whose resolution is provided in Section 8.5.

Table 8.1: Comparison of subcubes and Hamming balls or, equivalently, spheres

Regimes	Central limit		Moderate deviations	Large deviations
a, b	Fixed and large	Fixed and small	Subexp. vanishing	Exp. vanishing
Subcubes	Better	Worse	Worse	Worse
Balls/Spheres	Worse	Better	Better	Better

shells in Fig. 8.5. By the homogeneity property in (8.32) and (8.33), the surfaces corresponding to $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$ are formed by an infinite number of half-lines from the origin to infinity. Furthermore, by the relation between the MD, LD and CL exponents in (8.38)–(8.39) and (8.40)–(8.41), the surfaces of $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$ can be recovered from the surfaces of $\underline{\Upsilon}_{\mathrm{CL}}$ and $\overline{\Upsilon}_{\mathrm{CL}}$ by zooming them out, or recovered from $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ by zooming into a neighborhood of the origin. However, the surfaces of $\underline{\Upsilon}_{\mathrm{CL}}$ and $\overline{\Upsilon}_{\mathrm{CL}}$ as well as the surfaces of $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ *cannot* be recovered from those of $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$. In other words, $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$ contain much less information compared to $\underline{\Upsilon}_{\mathrm{CL}}$ and $\overline{\Upsilon}_{\mathrm{CL}}$ as well as $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$. This is not unexpected as the MD regime can be thought of a limiting case of the LD and CL regimes. Numerical results in Fig. 8.5 suggest that $\underline{\Upsilon}_{\mathrm{MD}}$ and $\underline{\Upsilon}_{\mathrm{LD}}$ are convex, and $\overline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ are concave, but $\underline{\Upsilon}_{\mathrm{CL}}$ and $\overline{\Upsilon}_{\mathrm{CL}}$ are neither convex nor concave. In Section 8.5, we discuss these issues rigorously in the context of the OPS conjecture (Conjecture 8.1).

We now compare the performances of subcubes, Hamming balls, and Hamming spheres (or spherical shells). We illustrate the forward joint probabilities achieved by subcubes and Hamming balls in Fig. 8.6. As the gaps between the probabilities are visually imperceptible, we also illustrate their differences on the right plot of Fig. 8.6. Based on the numerical comparisons, we observe that for large a and b, subcubes are better. However, for small a and b, Hamming balls are better. We summarize the performances of various geometric structures under different asymptotic regimes in Table 8.1. Based on these results, it is natural to ask whether subcubes are *optimal* for large a and b, and whether Hamming balls or spheres are *optimal* for small a and b. In the following sections, we provide answers to these questions.

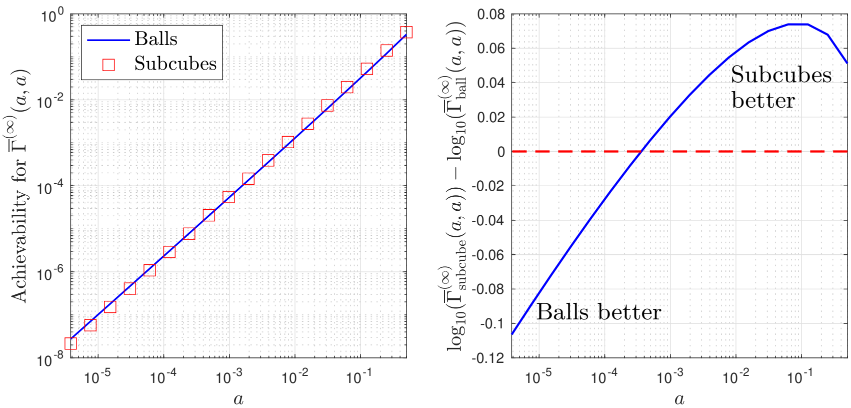

Figure 8.6: Left: The forward joint probabilities achieved by subcubes and Hamming balls with $a = b$ and $\rho = 0.5$; Right: The difference between the logarithms of the forward joint probabilities achieved by subcubes and Hamming balls which shows that subcubes outperform balls for large a and vice versa.

8.3 Converses in the Central Limit Regime

In this and the next two sections, we discuss the optimality of subcubes, Hamming balls, and spheres (or spherical shells) in the various asymptotic regimes for the forward and reverse joint probabilities. In this section, we consider the CL regime in which we are interested in determining whether subcubes are optimal in for the NICD problem for $a = b \in \{1/2, 1/4\}$. The case $a = b = 1/2$ is relatively well known and solved by Witsenhausen [178]. The case $a = b = 1/4$, however, is more challenging and, in fact, was posed as an open problem by E. Mossel in 2017 [119]; see also Mossel [120, Problem 2.6]. Here, we term the case $a = b = 1/4$, as the *mean*-$1/4$ *stability problem*. In the CL regime, it is also natural to ask whether Hamming balls are optimal for small but fixed a and b (i.e., $0 < a, b < 1/4$). Since this case behaves similarly to that in the MD regime, we will discuss it in the next section concerning the MD regime.

8.3.1 Case of $a = b = 1/2$: Maximal Correlation Method

We first consider the optimality of subcubes (or Boolean functions) for the case $a = b = 1/2$ in the NICD problem. By using the properties of the *maximal correlation*, the non-asymptotic optimality of subcubes for this basic case was confirmed positively by Witsenhausen [178]. We recall from (1.2) in the introduction that the *Hirschfeld–Gebelein–Rényi* (or *HGR*) *maximal correlation* [63], [84], [144] between two random variables X and Y is defined as

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

where $\rho(U;V)$ denotes the correlation coefficient between U and V (defined in (1.1)), and the supremum is taken over all real-valued functions f and g such that $0 < \mathrm{Var}(f(X)), \mathrm{Var}(g(Y)) < \infty$. It is well-known that the maximal correlation satisfies several desirable properties, including tensorization and the data processing inequality.

1. Tensorization: For a sequence of independent pairs of random variables $(X^n, Y^n) = \{(X_i, Y_i)\}_{i=1}^n$, we have

$$\rho_{\mathrm{m}}(X^n;Y^n) = \max_{i\in[n]} \rho_{\mathrm{m}}(X_i;Y_i). \tag{8.42}$$

2. Data processing inequality (DPI): For the Markov chain $U - X - Y - V$, we have

$$\rho_{\mathrm{m}}(U;V) \le \rho_{\mathrm{m}}(X;Y). \tag{8.43}$$

3. Binary random variables: For binary X and Y, we have

$$\rho_{\mathrm{m}}(X;Y) = |\rho(X;Y)|. \tag{8.44}$$

Using these properties, Witsenhausen [178] proved the following theorem.

Theorem 8.3. Let π_{XY} be the doubly symmetric binary distribution with correlation coefficient ρ as defined in (8.1). For any $\mathcal{A}$ and $\mathcal{B}$ with $\pi_X^n(\mathcal{A}) = a$ and $\pi_Y^n(\mathcal{B}) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le \pi_{XY}^n(\mathcal{A}\times\mathcal{B}) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}. \tag{8.45}$$

Proof. Let $(X^n, Y^n) \sim \pi_{XY}^n$. Define $U := \mathbb{1}_{\mathcal{A}}(X^n)$ and $V := \mathbb{1}_{\mathcal{B}}(Y^n)$. Then we have the Markov chain $U - X^n - Y^n - V$. Consider,

$$\begin{aligned}\frac{|\pi_{XY}^n(\mathcal{A}\times\mathcal{B}) - ab|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} &= |\rho(U;V)| \\ &= \rho_{\mathrm{m}}(U;V) \quad (8.46)\\ &\le \rho_{\mathrm{m}}(X^n;Y^n) \quad (8.47)\\ &= \rho_{\mathrm{m}}(X_1;Y_1) \quad (8.48)\\ &= \rho, \quad (8.49)\end{aligned}$$

where (8.46) and (8.49) follow from (8.44), (8.47) follows from the data processing inequality in (8.43), and (8.48) follows from the tensorization property in (8.42) (since all pairs of random variables are *identically distributed*, the max in (8.42) is simply $\rho_{\mathrm{m}}(X_1;Y_1)$). □

From Theorem 8.3, one deduces that for $a = b = 1/2$,

$$\frac{1-\rho}{4} \le \pi_{XY}^n(\mathcal{A}\times\mathcal{B}) \le \frac{1+\rho}{4}. \quad (8.50)$$

Based on the discussion around (8.13)–(8.14), the upper bound is achieved by a pair of identical dictator functions, i.e., $f(x^n) = g(x^n) = x_i$ (or $1 - x_i$) for all $i \in [n]$. Moreover, the lower bound is achieved by a pair anti-symmetric dictator functions, i.e., $f(x^n) = 1 - g(x^n) = x_i$ for all $i \in [n]$. Hence,

$$\overline{\Gamma}^{(n)}\Big(\frac{1}{2},\frac{1}{2}\Big) = \frac{1+\rho}{4} \quad \text{and} \quad \underline{\Gamma}^{(n)}\Big(\frac{1}{2},\frac{1}{2}\Big) = \frac{1-\rho}{4} \quad \text{for all } n \ge 1.$$

This result also can be proven by the hypercontractivity method and Fourier analysis; these are discussed in the next two subsections.

8.3.2 Case of $a = b = 1/2$: Hypercontractivity Method

The classic *hypercontractivity inequalities* form an important class of functional inequalities. These inequalities play a fundamental role in the NICD problem when the means of the Boolean functions are assumed to be either large or small. The forward and reverse parts of the hypercontractivity inequalities for the DSBS are stated in Theorem 8.4 which follow from Gross [69], Borell [27], and O'Donnell [131].

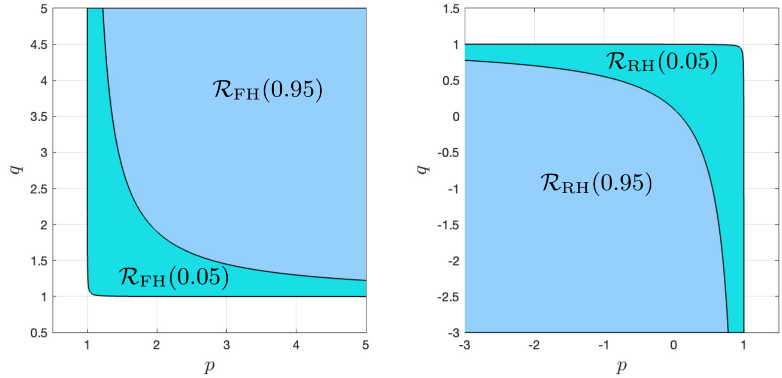

Figure 8.7: Plots of the forward (left) and reverse (right) hypercontractivity regions in (8.51) and (8.52) for $\rho = 0.05$ and 0.95

We commence with some definitions. For $f : \mathcal{X}^n \to [0, \infty)$ and $g : \mathcal{Y}^n \to [0, \infty)$, denote their *inner product*

$$\langle f, g \rangle := \mathbb{E}[f(X^n)g(Y^n)],$$

where the expectation is taken with respect to π_{XY}^n. Define the L^p*-norm* for $p \in [1, \infty)$ and the *pseudo* L^p*-norm* for $p \in (-\infty, 1)\backslash\{0\}$ as

$$\|f\|_p := (\mathbb{E}[f^p(X^n)])^{1/p}.$$

For $p \in \{0, \pm\infty\}$, $\|f\|_p$ is defined by its continuous extensions. Specifically,

$$\begin{aligned} \|f\|_0 &:= \mathrm{e}^{\mathbb{E}[\ln f(X^n)]}, \\ \|f\|_\infty &:= \max_{x^n \in \mathcal{X}^n} f(x^n), \quad \text{and} \\ \|f\|_{-\infty} &:= \min_{x^n \in \mathcal{X}^n} f(x^n), \end{aligned}$$

where $\|f\|_0$ is known as the *geometric mean* of f. Note that $\|f\|_p = 0$ for $p < 0$ if f is not positive π_X-almost everywhere.

For the DSBS $(X, Y) \sim \pi_{XY}$ with correlation coefficient ρ, define

$$\mathcal{R}_{\mathrm{FH}}(\rho) := \{(p, q) \in [1, \infty]^2 : (p-1)(q-1) \geq \rho^2\}, \text{ and} \tag{8.51}$$

$$\mathcal{R}_{\mathrm{RH}}(\rho) := \{(p, q) \in [-\infty, 1]^2 : (p-1)(q-1) \geq \rho^2\}. \tag{8.52}$$

These regions are respectively called the *forward* and *reverse hypercontractivity regions* for the DSBS and are illustrated in Fig. 8.7.

Theorem 8.4 (Hypercontractivity: DSBS and Two-Function Version)**.** Let $(X^n, Y^n) \sim \pi_{XY}^n$ be a source sequence generated by a DSBS with correlation coefficient ρ.

1. The inequality
$$\langle f, g\rangle \le \|f\|_p \|g\|_q \tag{8.53}$$
holds for all $f : \{0,1\}^n \to [0,\infty)$ and $g : \{0,1\}^n \to [0,\infty)$, if and only if $(p,q) \in \mathcal{R}_{\mathrm{FH}}(\rho)$.

2. The inequality
$$\langle f, g\rangle \ge \|f\|_p \|g\|_q \tag{8.54}$$
holds for all $f : \{0,1\}^n \to [0,\infty)$ and $g : \{0,1\}^n \to [0,\infty)$, if and only if $(p,q) \in \mathcal{R}_{\mathrm{RH}}(\rho)$.

These two inequalities (due to [27], [69], [131]) are known as the *two-function versions* of the hypercontractivity inequalities for the DSBS. These inequalities are equivalent to the following *single-function versions* of the hypercontractivity inequalities for the DSBS.

Before we describe these single-function versions, we introduce some additional notation. Denote $q' = \frac{q}{q-1}$ as the *Hölder conjugate* of q for $q \neq 1$; for $q = 1$, both $q = \pm\infty$ are Hölder conjugates of q. For a DSBS sequence $(X^n, Y^n) \sim \pi_{XY}^n = \pi_{X|Y}^n \times \pi_Y^n$ with correlation coefficient ρ, the *noise operator* or *conditional expectation operator* T_ρ (or $\pi_{X|Y}^n$) as

$$T_\rho f(y^n) := \mathbb{E}[f(X^n) \mid Y^n = y^n] = \sum_{x^n \in \mathcal{X}^n} f(x^n)\pi_{X|Y}^n(x^n|y^n). \tag{8.55}$$

One can easily check that $T_{\rho_1\rho_2} = T_{\rho_1}T_{\rho_2}$ for all $\rho_1, \rho_2 \in [0,1]$.

Theorem 8.5 (Hypercontractivity: DSBS and Single-Function Version)**.** Let $(X^n, Y^n) \sim \pi_{XY}^n$ be a source sequence generated by a DSBS with correlation coefficient ρ.

1. The inequality
$$\|T_\rho f\|_q \le \|f\|_p \tag{8.56}$$
holds for all $f : \{0,1\}^n \to [0,\infty)$, if and only if $(p, q') \in \mathcal{R}_{\mathrm{FH}}(\rho)$ (with $1' := \infty$).

2. The inequality

$$\|T_\rho f\|_q \geq \|f\|_p \tag{8.57}$$

holds for all $f : \{0,1\}^n \to [0,\infty)$, if and only if $(p, q') \in \mathcal{R}_{\mathrm{RH}}(\rho)$ (with $1' := -\infty$).

Here we do not delve deeper into the equivalence between the single- and two-function versions of hypercontractivity inequalities, since we will discuss the equivalence in detail in Section 10.2.3.

By applying the hypercontractivity inequalities, Kamath and Anantharam [94, Eqns. (28) and (29)] provided the following bounds.

Theorem 8.6 (Hypercontractivity bound for the DSBS)**.** Define the function

$$\varphi_{a,b}(s,t,p) := \frac{(s^p a + \overline{a})^{\frac{1}{p}} (t^q b + \overline{b})^{\frac{1}{q}} - 1}{(s-1)(t-1)} - \frac{a}{t-1} - \frac{b}{s-1}$$

with $q := 1 + \rho^2/(p-1)$. Then, for any sets $\mathcal{A}$ and $\mathcal{B}$ with $\pi_X^n(\mathcal{A}) = a$ and $\pi_Y^n(\mathcal{B}) = b$,

$$\sup_{s,t>0, p:(s-1)(t-1)(p-1)<0} \varphi_{a,b}(s,t,p) \leq \pi_{XY}^n(\mathcal{A} \times \mathcal{B}) \tag{8.58}$$

$$\leq \inf_{s,t>0, p:(s-1)(t-1)(p-1)>0} \varphi_{a,b}(s,t,p). \tag{8.59}$$

Proof. This theorem follows by setting f and g in Theorem 8.4 to be $\{s,1\}$-valued and $\{t,1\}$-valued functions respectively. Note that changing the range of the functions f and g from $\{0,1\}$ to the sets $\{s,1\}$ and $\{t,1\}$ respectively does not affect the values of the probability masses of the joint distribution of $(f(X^n), g(Y^n))$. $\square$

It can be shown analytically that the hypercontractivity bounds are no worse than the maximal correlation bounds in Theorem 8.3 for any $a, b \in [0,1]$; see Fig. 8.8 for a numerical comparison. Moreover, for $a = b = 1/2$, the hypercontractivity bounds in (8.58) and (8.59) reduce to the sharp bounds $\frac{1-\rho}{4} \leq \pi_{XY}^n(\mathcal{A} \times \mathcal{B}) \leq \frac{1+\rho}{4}$, which correspond to the bounds given by the maximal correlation technique in (8.50).

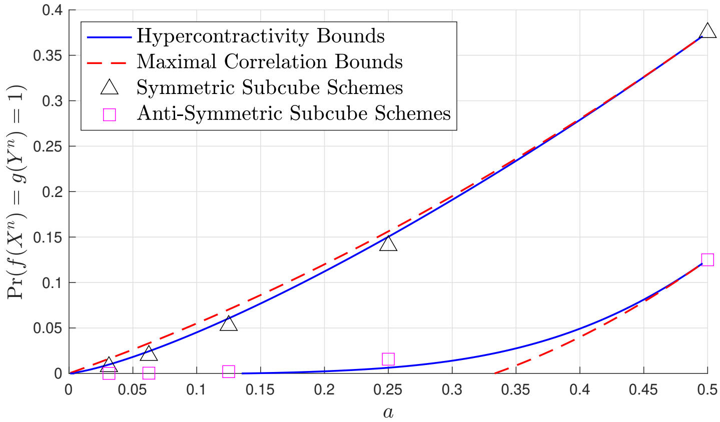

Figure 8.8: Illustration the maximal correlation bounds in (8.45), the hypercontractivity bounds in (8.58)–(8.59) as well as the performances of symmetric and anti-symmetric subcube schemes

8.3.3 Case of $a = b = 1/4$: Boolean Fourier Analysis

We now consider the case $a = b = 1/4$, and we answer the forward part of Mossel's mean-1/4 stability problem. Mossel's mean-1/4 stability problem [119], [120] consists in the determination of $\overline{\Gamma}^{(n)}(1/4, 1/4)$ (forward part) and $\underline{\Gamma}^{(n)}(1/4, 1/4)$ (reverse part) for $n \geq 2$, and also the optimal Boolean functions that attain the maximum and minimum that define these two quantities.

The forward part of this problem was resolved by the present authors in [198], [205] using elements of *Boolean Fourier analysis*. We recap some fundamentals of this study here. Given a Boolean function $f : \{0,1\}^n \to \{0,1\}$, its *Fourier coefficients* are defined as

$$\hat{f}_{\mathcal{S}} := \mathbb{E}[f(X^n)\chi_{\mathcal{S}}(X^n)] = \frac{1}{2^n} \sum_{x^n \in \{0,1\}^n} f(x^n)\, \chi_{\mathcal{S}}(x^n) \quad \text{for all } \mathcal{S} \subset [n],$$

where the *(Fourier) basis functions* are

$$\chi_{\mathcal{S}}(x^n) := (-1)^{\sum_{i\in\mathcal{S}} x_i} \quad \text{for all } x^n \in \{0,1\}^n,$$

and $X^n \sim \mathrm{Unif}\{0,1\}^n$. The function f can be expressed in terms of the

Fourier coefficients as

$$f(x^n) = \sum_{\mathcal{S}\subset[n]} \hat{f}_{\mathcal{S}}\, \chi_{\mathcal{S}}(x^n) \quad \text{for all } x^n \in \{0,1\}^n,$$

which is known as the *Fourier expansion* of f. For $0 \le k \le n$, define the *degree-k Fourier weight* of f as

$$\mathbf{W}_k[f] := \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|=k} \hat{f}_{\mathcal{S}}^2. \tag{8.60}$$

It is easy to check that if we define the *degree-k part of f* as $f_k(x^n) := \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|=k} \hat{f}_{\mathcal{S}}\, \chi_{\mathcal{S}}(x^n)$, then $\mathbb{E}[f_k(X^n)^2] = \mathbf{W}_k[f]$. Hence, $\mathbf{W}_k[f]$ represents the "energy" of the degree-k part in f's Fourier expansion.

The Fourier weights satisfy the following properties. Proofs of these properties can be found in the delightful exposition of Boolean functions by O'Donnell [131].

Lemma 8.7. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ with mean a,

$$\mathbf{W}_0[f] = a^2 \quad \text{and} \quad \sum_{k=0}^{n} \mathbf{W}_k[f] = a. \tag{8.61}$$

Furthermore, if $(X^n, Y^n) \sim \pi_{XY}^n$ is a source sequence of the DSBS with correlation coefficient ρ, then for any pair of Boolean functions $f, g : \{0,1\}^n \to \{0,1\}$,

$$\Pr(f(X^n) = g(Y^n) = 1) = \sum_{k=0}^{n} \rho^k \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|=k} \hat{f}_{\mathcal{S}}\, \hat{g}_{\mathcal{S}} \quad \text{and}$$

$$\Pr(f(X^n) = f(Y^n) = 1) = \sum_{k=0}^{n} \mathbf{W}_k[f]\, \rho^k.$$

For $\rho \in (0,1)$, lower degree Fourier weights have a higher contribution to the joint probability $\Pr(f(X^n) = f(Y^n) = 1)$ than higher degree weights. Hence, to bound this joint probability, we can focus on bounding the lower degree Fourier weights of f. Observe from (8.61) that given the mean of f, the degree-0 Fourier weight is fully specified. Hence, it is instructive to estimate the *second most important* Fourier weight. In particular, we are interested in the *degree-*1 *Fourier weight* $\mathbf{W}_1[f]$ under the condition that the mean of f is specified. In the literature, there exist several bounds on $\mathbf{W}_1[f]$. These include Chang's bound, which can be found in [131, Level-1 Inequality] and [35] and the linear programming (LP) bounds of Fu, Wei, and Yeung [59] and Yu and Tan [198]. In particular, the LP bounds state that

$$\mathbf{W}_1[f] \le \varphi(a) := \begin{cases} 2a(\sqrt{a} - a) & 0 \le a \le 1/4 \\ a/2 & 1/4 < a \le 1/2 \end{cases}. \tag{8.62}$$

By the Cauchy–Schwarz inequality, one easily observes that

$$\begin{aligned} &\Pr(f(X^n) = g(Y^n) = 1) \\ &\quad \le \max\left\{\Pr(f(X^n)=f(Y^n)=1), \Pr(g(X^n)=g(Y^n)=1)\right\}. \end{aligned} \tag{8.63}$$

This inequality implies that in the determination of $\overline{\Gamma}^{(n)}(a, a)$ (the symmetric case in which $a = b$), it suffices to consider a pair of *identical* Boolean functions.

By combining the ideas in Lemma 8.7, the LP bound in (8.62) and (8.63), the present authors proved the following result [198], [205].

Theorem 8.8. For all $a \in [0, 1]$ and $n \ge 2$,

$$\overline{\Gamma}^{(n)}(a, a) \le a^2 + \rho\varphi(a) + \rho^2(a - a^2 - \varphi(a)).$$

Particularizing this upper bound to $a = b = 1/4$, we obtain $\overline{\Gamma}^{(n)}(1/4, 1/4) \le (\frac{1+\rho}{4})^2$. Per the discussion leading to (8.13), this upper bound is attained by a pair of identical $(n-2)$-subcubes. Hence,

$$\overline{\Gamma}^{(n)}\Big(\frac{1}{4}, \frac{1}{4}\Big) = \Big(\frac{1+\rho}{4}\Big)^2 \quad \text{for all } n \ge 2,$$

resolving the forward part of Mossel's mean-1/4 stability problem. However, the reverse part of the same problem (i.e., which Boolean functions attain $\underline{\Gamma}^{(n)}(1/4, 1/4)$) remains open.

8.4 Converse in the Moderate Deviations Regime

We now consider the optimality of Hamming balls and spheres in the MD regime and the CL regime with small a and b. To address this question, we resort to two key ideas, namely the hypercontractivity inequalities in Theorem 8.4 and the *small set expansion (SSE)* theorem.

A well-known result to address the optimality of Hamming balls and spheres in the MD regime and the CL regime with small a and b is the SSE theorem [124], [131], which is a consequence of the hypercontractivity inequalities in Theorem 8.4.

Theorem 8.9 (Small set expansion: DSBS version)**.** For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha,\beta) \geq \underline{\Upsilon}_{\mathrm{MD}}(\alpha,\beta) \quad \text{and}$$
$$\overline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha,\beta) \leq \overline{\Upsilon}_{\mathrm{MD}}(\alpha,\beta),$$

where $\underline{\Upsilon}_{\mathrm{MD}}$ and $\overline{\Upsilon}_{\mathrm{MD}}$ are expressed in closed form for the DSBS in (8.34) and (8.35) respectively.

The reader might wonder about the term "small set expansion" that is used to describe Theorem 8.9. This term refers to a curious phenomenon of the Hamming cube being a "small set expander" in the sense that any small subset $\mathcal{A} \subset \{0,1\}^n$ has an usually large (or expanded) boundary. Here, the Hamming cube is regarded as an edge-weighted complete graph, known as the *ρ-stable hypercube graph*, in which each edge (x^n, y^n) is assigned a weight equal to the probability $\pi_{XY}^n(x^n, y^n)$. The limiting case as $\rho \downarrow 0$ of this phenomenon is quantified by the edge-isoperimetric inequality which will be stated in Theorem 9.5. We refer readers to O'Donnell [131] for more intuition about the term "small set expansion".

Proof Sketch of Theorem 8.9. Substituting the indicator functions $f \leftarrow \mathbb{1}_{\mathcal{A}}$ and $g \leftarrow \mathbb{1}_{\mathcal{B}}$ into (8.53) and (8.54) respectively, and optimizing over (p,q), we obtain the inequalities as stated in the SSE theorem. $\square$

Due to the equivalence among the CL, MD, and LD exponents for all $n \in \mathbb{N}$ (as discussed after Definition 8.2) and the homogeneity property in (8.32) and (8.33), $\underline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha,\beta)$ and $\overline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha,\beta)$ in Theorem 8.9 can be replaced by $\underline{\Upsilon}_{\mathrm{CL}}^{(n)}(\alpha,\beta)$ and $\overline{\Upsilon}_{\mathrm{CL}}^{(n)}(\alpha,\beta)$ respectively, or by $\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha,\beta)$ and $\overline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha,\beta)$ respectively.

The bounds in the SSE theorem are achieved by sequences of Hamming balls or spherical shells. Hence, these geometric objects are optimal in attaining the MD exponents.

8.5 Converse in the Large Deviations Regime

We now address the final asymptotic regime of interest, namely, the large deviations regime. First, we introduce some terminology. Let $\mathcal{I} \subset \mathbb{R}^d$ be a convex subset of d-dimensional Euclidean space. We recall that for a function $f : \mathcal{I} \to \mathbb{R}$, its *lower convex envelope* $\mathbb{L}[f]$ is the function defined at each point of $\mathcal{I}$ as the supremum of all convex functions that lie under f, i.e., for every $\mathbf{x} \in \mathbb{R}^d$, $\mathbb{L}[f](\mathbf{x}) := \sup\{g(\mathbf{x}) : g \text{ is convex}, g \leq f \text{ on } \mathcal{I}\}$. By Carathéodory's theorem, equivalently,

$$\mathbb{L}[f](\mathbf{x}) = \inf_{\{\mathbf{x}_i\}_{i=1}^{d+1} \subset \mathcal{I},\ \{\lambda_i\}_{i=1}^{d+1}} \sum_{i=1}^{d+1} \lambda_i f(\mathbf{x}_i), \tag{8.64}$$

where $\{\lambda_i\}_{i=1}^{d+1}$ is a $(d+1)$-dimensional probability mass function with $\sum_{i=1}^{d+1} \lambda_i \mathbf{x}_i = \mathbf{x}$. The *upper concave envelope* $\mathbb{U}[f]$ is defined as $-\mathbb{L}[-f]$. The SSE theorem in Theorem 8.9 can be strengthened to the following result, known as the *strong SSE theorem*; see Yu, Anantharam, and Chen [195] and Yu [192].

Theorem 8.10 (Strong small set expansion: DSBS version)**.** For any $n \geq 1$ and $\alpha, \beta \in (0, 1]$,

$$\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \mathbb{L}[\underline{\Upsilon}_{\mathrm{LD}}](\alpha, \beta) \quad \text{and} \tag{8.65}$$

$$\overline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \mathbb{U}[\overline{\Upsilon}_{\mathrm{LD}}](\alpha, \beta). \tag{8.66}$$

The proof of this theorem (and also its generalization to the finite alphabet case in Theorem 8.11) will be provided in Section 10.3. The proof is based on the information-theoretic characterizations of hypercontractivity inequalities (also discussed in Section 10).

By Carathéodory's representation of the lower convex and upper concave envelopes in (8.64), the bounds in Theorem 8.10 can be asymptotically achieved by "time-sharing" at most three (since $d = 2$ in our case) concentric or anti-concentric Hamming spheres (or balls) for each length n. Specifically, let $(\lambda_1, \lambda_2, \lambda_3)$ be a PMF, i.e., $\lambda_i \geq 0$ for all $i \in [3]$ and $\sum_{i=1}^{3} \lambda_i = 1$. For each blocklength $n \in \mathbb{N}$, this strategy uses certain concentric or anti-concentric Hamming spheres $\mathbb{S}^{(i)}$ for a period of length $\lfloor n\lambda_i \rfloor, i \in [3]$. Since time-sharing of certain Hamming spheres is optimal in the LD regime, this confirms a weaker version of the OPS conjecture (Conjecture 8.1) in which the convexification and concavification operations in (8.65) and (8.66) respectively are permitted.

Theorem 8.10 is known as the *strong* SSE theorem because the bounds given in Theorem 8.10 are asymptotically sharp in the LD regime. This is in contrast to the ones given in the vanilla SSE theorem (Theorem 8.9) which are not sharp in the LD regime. Furthermore, both these two theorems are asymptotically sharp in the MD regime, since the bounds in the strong SSE theorem reduce to the ones in the SSE theorem, as shown in (8.38) and (8.39). Hence, Theorem 8.10 is stronger than the SSE theorem (Theorem 8.9), in the sense that for all $\alpha, \beta \in [0,1]$ and $\gamma > 0$,

$$\begin{aligned}\mathbb{L}[\underline{\Upsilon}_{\mathrm{LD}}](\alpha,\beta) &\geq \underline{\Upsilon}_{\mathrm{MD}}(\gamma\alpha,\gamma\beta) \quad \text{and}\\ \mathbb{U}[\overline{\Upsilon}_{\mathrm{LD}}](\alpha,\beta) &\leq \overline{\Upsilon}_{\mathrm{MD}}(\gamma\alpha,\gamma\beta).\end{aligned}$$

To prove the OPS conjecture, we need to remove the operations of taking the lower convex and upper concave envelopes in the strong SSE theorem. This was done by the first author of this monograph [193]. In particular, he showed that $\underline{\Upsilon}_{\mathrm{LD}}$ is convex and $\overline{\Upsilon}_{\mathrm{LD}}$ is concave. Combining this result with the strong SSE theorem (Theorem 8.10) allows us to conclude that the OPS conjecture is unconditionally true and that Hamming balls or spheres (without time-sharing) are optimal in the LD regime [193]. That is, for the DSBS and $\alpha, \beta \in (0,1)$,

$$\underline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) = \underline{\Upsilon}_{\mathrm{LD}}(\alpha,\beta) \quad \text{and} \quad \overline{\Upsilon}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) = \overline{\Upsilon}_{\mathrm{LD}}(\alpha,\beta). \tag{8.67}$$

Several special cases of (8.67) were established in the literature prior to the most general result of Yu [193]. The limiting cases as $\rho \downarrow 0$ and $\rho \uparrow 1$ were shown by Ordentlich, Polyanskiy, and Shayevitz [133]. The "symmetric" special case with $\alpha = \beta$ was shown by Kirshner and Samorodnitsky [97]. We introduce these results in Section 10, since they are consequences of strengthened versions of the hypercontractivity inequalities.

We summarize all converse results discussed in Sections 8.3–8.5 and techniques used to prove them in Table 8.2.

Table 8.2: Converse (optimality) results and techniques for the 2-user NICD problem in the CL, MD, and LD regimes

<table>
<tr><th rowspan="2">Regimes</th><th colspan="2">Central Limit</th><th>Moderate Deviations</th><th>Large Deviations</th></tr>
<tr><th>Fixed and large a, b</th><th>Fixed but small a, b</th><th>Subexp. vanishing a, b</th><th>Exp. vanishing a, b</th></tr>
<tr><td>Maximal Correlation</td><td>Sharp for $a = b = 1/2$</td><td>Not sharp</td><td>Not sharp</td><td>Not sharp</td></tr>
<tr><td>Fourier Analysis</td><td>Sharp for $a = b = 1/2$ and $a = b = 1/4$</td><td>Not sharp</td><td>Not sharp</td><td>Not sharp</td></tr>
<tr><td>SSE</td><td>Not sharp</td><td rowspan="2">Essentially sharp</td><td rowspan="2">Sharp</td><td>Not sharp</td></tr>
<tr><td>Strong SSE</td><td>Not sharp</td><td>Sharp</td></tr>
</table>

8.6 Extensions to Sources Beyond the DSBS

Thus far, we have only considered the DSBS. Can the results in Sections 8.2-8.5 be extended to other bivariate memoryless sources? Indeed, the SSE and strong SSE theorems, can be extended to sources on Polish spaces (separable completely metrizable topological space). We refer the reader to [192] for details. Here for simplicity, we discuss analogues of the preceding results for the finite alphabet and bivariate Gaussian cases. The NICD problem for the latter case has been completely solved by Borell [28] and Mossel and Neeman [121].

8.6.1 Finite Alphabets

In this section, we generalize the NICD problem to the finite alphabet case in which $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. Let $\pi_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. For simplicity, we assume that the supports of π_X and π_Y are $\mathcal{X}$ and $\mathcal{Y}$ respectively. Given π_X and π_Y, define their maximum exponents of "atomic events" as

$$\begin{aligned} \alpha_{\max}(\pi_X) &:= \max_{x \in \mathcal{X}} \log \frac{1}{\pi_X(x)} \qquad \text{and} \\ \beta_{\max}(\pi_Y) &:= \max_{y \in \mathcal{Y}} \log \frac{1}{\pi_Y(y)}. \end{aligned} \tag{8.68}$$

For $n \geq 1$, $\alpha \in (0, \alpha_{\max}(\pi_X)]$ and $\beta \in (0, \beta_{\max}(\pi_Y)]$, re-define the *forward* and *reverse LD exponents* respectively as

$$\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \max_{\substack{\mathcal{A} \subset \mathcal{X}^n, \mathcal{B} \subset \mathcal{Y}^n: \\ \pi_X^n(\mathcal{A}) \leq 2^{-n\alpha}, \pi_Y^n(\mathcal{B}) \leq 2^{-n\beta}}} \pi_{XY}^n(\mathcal{A} \times \mathcal{B}) \quad \text{and} \tag{8.69}$$

$$\overline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \min_{\substack{\mathcal{A} \subset \mathcal{X}^n, \mathcal{B} \subset \mathcal{Y}^n: \\ \pi_X^n(\mathcal{A}) \geq 2^{-n\alpha}, \pi_Y^n(\mathcal{B}) \geq 2^{-n\beta}}} \pi_{XY}^n(\mathcal{A} \times \mathcal{B}). \tag{8.70}$$

Let $\underline{\Upsilon}_{\mathrm{LD}}^{(\infty)}$ and $\overline{\Upsilon}_{\mathrm{LD}}^{(\infty)}$ be their pointwise limits as $n \to \infty$. These are the same as the forward and reverse LD exponents in (8.9) and (8.10) but here, π_{XY} is no longer restricted to be a DSBS.

Theorem 8.11 (Strong small set expansion: General version)**.** For any joint distribution on a finite alphabet π_{XY}, any blocklength $n \geq 1$, $\alpha \in (0, \alpha_{\max}(\pi_X)]$, and $\beta \in (0, \beta_{\max}(\pi_Y)]$, (8.65) and (8.66) remain true, with $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ defined in (8.24) and (8.25) for π_{XY}, i.e.,

$$\underline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) = \min_{Q_X, Q_Y : D(Q_X \| \pi_X) \geq \alpha, D(Q_Y \| \pi_Y) \geq \beta} \mathsf{D}(Q_X, Q_Y \| \pi_{XY})$$

and analogously for $\overline{\Upsilon}_{\mathrm{LD}}$. Moreover, the inequalities in (8.65) and (8.66) remain asymptotically tight in the limit as $n \to \infty$.

However, in general, $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ are not necessarily convex and concave, respectively. Hence, unlike the case of the DSBS, for sources on finite alphabets, the operations of taking the lower convex and upper concave envelopes in (8.65) and (8.66) cannot be removed in general. Nevertheless, the bounds $\mathbb{L}[\underline{\Upsilon}_{\mathrm{LD}}](\alpha, \beta)$ and $\mathbb{U}[\overline{\Upsilon}_{\mathrm{LD}}](\alpha, \beta)$ can be asymptotically attained by *time-sharing* the use of at most three type classes (cf. the discussion after Theorem 8.10).

Theorem 8.11 was first proven by Yu, Anantharam, and Chen [195] by using information-theoretic and coupling techniques. In this monograph, we will provide a simple proof of Theorem 8.11, which is based on the information-theoretic characterizations of hypercontractivity inequalities as discussed in Section 10.3.

Similarly, one can generalize the DSBS-specific definitions in (8.11) and (8.12) to an arbitrary distribution π_{XY} on a finite alphabet. Then, the SSE theorem (Theorem 8.9) can be also generalized to the finite alphabet case.

Theorem 8.12 (Small set expansion: General version)**.** For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha, \beta) \geq \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbb{L}[\underline{\Upsilon}_{\mathrm{LD}}](\epsilon\alpha, \epsilon\beta) \quad \text{and} \tag{8.71}$$

$$\overline{\Upsilon}_{\mathrm{MD}}^{(n)}(\alpha, \beta) \leq \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbb{U}[\overline{\Upsilon}_{\mathrm{LD}}](\epsilon\alpha, \epsilon\beta). \tag{8.72}$$

Moreover, the inequalities in (8.71) and (8.72) are asymptotically tight in the limit as $n \to \infty$.

Since, in general, $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ are not necessarily convex and concave, respectively, the operations of taking the lower convex and upper concave envelopes in (8.71) and (8.72) cannot be removed as well. As a consequence, for this case, (8.71) and (8.72) cannot be written as in the variational expressions that appear on right-hand sides of (8.29) and (8.30).

8.6.2 Gaussian Sources

We next consider memoryless bivariate Gaussian sources with correlation coefficient $\rho \in (-1, 1) \setminus \{0\}$. For such sources, the NICD problem was completely solved by Borell [28] (for the symmetric cases in which $a = b$) and Mossel and Neeman [121] (for the asymmetric cases) for all $(a, b) \in [0, 1]^2$ and *non-asymptotically*, i.e., for all n. Let π_{XY} be the bivariate Gaussian distribution with mean $(0, 0)$ and covariance matrix $\mathbf{K}$ given in (8.17), where the correlation coefficient $\rho \in (-1, 1) \setminus \{0\}$. As usual, let $(X^n, Y^n) \sim \pi_{XY}^n$.

Theorem 8.13 (Borell's isoperimetric theorem)**.** For any $n \geq 1$ and $a, b \in [0, 1]$,

$$\overline{\Gamma}^{(n)}(a, b) = \Lambda_\rho(a, b) \quad \text{and} \quad \underline{\Gamma}^{(n)}(a, b) = \Lambda_{-\rho}(a, b),$$

where the bivariate normal copula $\Lambda_\rho(\cdot, \cdot)$ is defined in (8.19).

Moreover, it has been shown by Mossel and Neeman [121] that the optimal subsets $(\mathcal{A}, \mathcal{B})$ attaining $\overline{\Gamma}^{(n)}$ or $\underline{\Gamma}^{(n)}$ must be equal to parallel halfspaces (almost everywhere).

Specialized to the case of $a = b = 1/2$, this theorem implies that

$$\overline{\Gamma}^{(n)}\Big(\frac{1}{2},\frac{1}{2}\Big) = \frac{1}{2} - \frac{\arccos\rho}{2\pi} \quad \text{and} \quad \underline{\Gamma}^{(n)}\Big(\frac{1}{2},\frac{1}{2}\Big) = \frac{\arccos\rho}{2\pi}. \tag{8.73}$$

The optimal $(\mathcal{A},\mathcal{B})$ attaining $\overline{\Gamma}^{(n)}(1/2,1/2)$ correspond to a pair of identical halfspaces through the origin. In contrast, the optimal $(\mathcal{A},\mathcal{B})$ attaining $\underline{\Gamma}^{(n)}(1/2,1/2)$ correspond to a pair of complementary halfspaces through the origin.

Next, we provide a proof sketch of Theorem 8.13 which is due to Mossel and Neeman [121]. In fact, they also proved the following equivalent form of Theorem 8.13.

Theorem 8.14. For any $n \geq 1$, any pair of measurable functions $f, g : \mathbb{R}^n \to [0,1]$, and any $0 < \rho < 1$,

$$\mathbb{E}\big[\Lambda_\rho(f(X^n), g(Y^n))\big] \leq \Lambda_\rho\big(\mathbb{E}[f(X^n)], \mathbb{E}[g(Y^n)]\big). \tag{8.74}$$

If $-1 < \rho < 0$, the inequality in (8.74) is reversed.

To see that Theorem 8.14 implies Theorem 8.13, set $f = \mathbb{1}_{\mathcal{A}}$ and $g = \mathbb{1}_{\mathcal{B}}$ for two sets $\mathcal{A},\mathcal{B} \subset \mathbb{R}^n$ such that $\mathbb{E}[f(X^n)] = a$ and $\mathbb{E}[g(Y^n)] = b$ in Theorem 8.14. Observe that $\Lambda_\rho(0,0) = \Lambda_\rho(1,0) = \Lambda_\rho(0,1) = 0$, and $\Lambda_\rho(1,1) = 1$. Therefore, $\Lambda_\rho(f(X^n), g(Y^n)) = \mathbb{1}_{\mathcal{A}\times\mathcal{B}}(X^n, Y^n)$, which implies that $\overline{\Gamma}^{(n)}(a,b) \leq \Lambda_\rho(a,b)$. Obviously, by definition, $\overline{\Gamma}^{(n)}(a,b) \geq \Lambda_\rho(a,b)$ follows by setting $\mathcal{A}$ and $\mathcal{B}$ to be two parallel halfspaces. Hence, $\overline{\Gamma}^{(n)}(a,b) = \Lambda_\rho(a,b)$.

We now argue that Theorem 8.13 implies Theorem 8.14. For this purpose, given $f, g : \mathbb{R}^n \to [0,1]$, define $\mathcal{A}$ and $\mathcal{B}$ (subsets of $\mathbb{R}^{n+1}$) to be the respective hypographs[1] of $\Phi^{-1} \circ f : \mathbb{R}^n \to \mathbb{R}$ and $\Phi^{-1} \circ g : \mathbb{R}^n \to \mathbb{R}$, where recall that $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian and $\Phi^{-1} : (0,1) \to \mathbb{R}$ is its inverse. It can be readily checked that

$$\begin{aligned}\mathbb{E}\big[\Lambda_\rho(f(X^n), g(Y^n))\big] &= \Pr\left(X_{n+1} \leq \Phi^{-1} \circ f(X^n), Y_{n+1} \leq \Phi^{-1} \circ g(Y^n)\right)\\ &= \pi_{XY}^{n+1}(\mathcal{A}\times\mathcal{B}),\end{aligned}$$

[1] The *hypograph* $\mathrm{hyp}(h)$ of a function $h : \mathcal{X} \to \mathbb{R}$ is the set of points of $\mathcal{X} \times \mathbb{R}$ lying on or below its graph, i.e., $\mathrm{hyp}(h) := \{(x,r) \in \mathcal{X} \times \mathbb{R} : r \leq h(x)\}$.

where $(X^{n+1}, Y^{n+1}) \sim \pi_{XY}^{n+1}$. On the other hand, $\mathbb{E}[f(X^n)] = \pi_X^{n+1}(\mathcal{A})$ and $\mathbb{E}[g(Y^n)] = \pi_Y^{n+1}(\mathcal{B})$, and hence, the right-hand side of (8.74) satisfies

$$\Lambda_\rho(\mathbb{E}[f(X^n)], \mathbb{E}[g(Y^n)]) = \Lambda_\rho(\pi_X^{n+1}(\mathcal{A}), \pi_Y^{n+1}(\mathcal{B})).$$

Thus, Theorem 8.13 in $n+1$ dimensions implies Theorem 8.14 in n dimensions.

Hence, to prove Theorem 8.13, it suffices to prove Theorem 8.14. In their proof of Theorem 8.14, Mossel and Neeman [121] first constructed an Ornstein–Uhlenbeck semigroup, then defined R_t, an auxiliary function for this semigroup that connects the two sides of (8.74) as limiting cases. Lastly, they showed that R_t is monotone. A similar idea was also used in Bakry and Ledoux [10].

Proof Sketch of Theorem 8.14. For every $t \geq 0$, define the operator P_t that acts on functions $f : \mathbb{R}^n \to [0, 1]$ as

$$(P_t f)(x^n) := \int_{\mathbb{R}^n} f\big(\mathrm{e}^{-t}\, x^n + \sqrt{1 - \mathrm{e}^{-2t}}\, y^n\big)\, \mathrm{d}\pi_Y^n(y^n).$$

This operator is known as the *Ornstein–Uhlenbeck semigroup operator*. Note that $P_t f \to f$ pointwise as $t \to 0$ and $P_t f \to \mathbb{E}[f]$ pointwise as $t \to \infty$.

Let $f_t := P_t f$ and $g_t := P_t g$, and consider the quantity

$$R_t := \mathbb{E}\big[\Lambda_\rho(f_t(X^n), g_t(Y^n))\big]. \tag{8.75}$$

As $t \to 0$, R_t converges to the left-hand side of (8.74); as $t \to \infty$, R_t converges to the right-hand side of (8.74). Hence, to establish Theorem 8.14, it suffices to prove that $\mathrm{d}R_t/\mathrm{d}t \geq 0$ for all $t > 0$. This point can be checked by careful calculations, as shown in the following lemma due to Mossel and Neeman [121].

Lemma 8.15. The function $t \in [0, \infty) \mapsto R_t$, defined in (8.75), satisfies

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = \frac{\rho}{2\pi\sqrt{1-\rho^2}}\, \mathbb{E}\bigg[\exp\bigg(-\frac{v_t^2 + w_t^2 - 2\rho v_t w_t}{2(1-\rho^2)}\bigg)\bigg]\ \|\nabla v_t - \nabla w_t\|^2,$$

where $v_t := \Phi^{-1} \circ f_t : \mathbb{R}^n \to \mathbb{R}$, $w_t := \Phi^{-1} \circ g_t : \mathbb{R}^n \to \mathbb{R}$, and ∇ denotes the gradient operator. Hence, the derivative of R_t for $t \geq 0$ is nonnegative.

This completes the proof sketch of Theorem 8.14. ☐

9

q-Stability

In Section 8, we discussed the 2-user NICD problem. In this section, we extend the NICD problem to the multi-user case, and consider two versions of these extensions. In the *symmetric* version, we maximize the agreement probability of the random bits generated individually by the users. In the *asymmetric* version, we maximize the joint probability that all the random bits are equal to 1. These two maximization problems are equivalent in the 2-user setting (see discussion following (8.3)), but are *not* equivalent in the setting involving 3 or more users. This distinction results in the upcoming set of problems being significantly more challenging, but they provide more insight into the NICD and related problems.

Indeed, these extensions have inspired researchers to define a more general concept known as the *q-stability*. This is done by generalizing the number of users in the NICD problem from an integer k to an arbitrary real number $q \geq 1$. The *max q-stability* problem concerns the identification of Boolean functions that most "stable"—measured in terms of the q-stability—under the action of a noise operator. Such a problem not only significantly generalizes the 2-user NICD problem to a version parametrized by an arbitrary real number $q \geq 1$, but more importantly, it seamlessly connects to several interesting contemporary conjectures in information theory and discrete probability, including the Mossel–O'Donnell conjecture [122], the Courtade–Kumar conjecture [40], and the Li–Médard conjecture [110]. Hence, the study of q-stability provides us a comprehensive and unified understanding of these conjectures.

Similar to Section 8, in this section, we focus mainly on the doubly symmetric binary source (DSBS) with correlation coefficient $\rho \in (-1, 1)$. In Section 9.1, we formulate the multi-user NICD problem for the DSBS. We define the asymmetric and symmetric forward joint probabilities, and also generalize them to various max q-stabilities by relaxing the number of users to an arbitrary real number $q \geq 1$. In Section 9.2, we introduce several important conjectures concerning the max q-stability problem for the case in that the Boolean functions in question are *balanced*. These include the Mossel–O'Donnell, Courtade–Kumar, and Li–Médard conjectures. In Section 9.3, we describe resolutions for the conjectures in the extreme cases in which the correlation coefficient $\rho \downarrow 0$ or $\rho \uparrow 1$. Interestingly, in these two extreme cases, the conjectures are characterized by the classic *edge-isoperimetric inequality* and the *maximal degree-*1 *Fourier weight*. Hence, related concepts in discrete geometry, e.g., influences and edge boundaries, will also be introduced. In Section 9.4, we describe recent progress on partial resolutions of these conjectures. In Section 9.5, we introduce the solutions to the max q-stability problem in the moderate and large deviations regimes. Finally, in Section 9.6, we discuss known results on the max q-stability problem for sources beyond the DSBS including bivariate Gaussian sources.

9.1 The Multi-User NICD Problem and q-Stability

9.1.1 Formulation

Before formally introducing the k-user NICD problem, we first introduce a class of Boolean functions, known as *majority functions*. For an odd number $m \in [n]$, let $\mathrm{Maj}_m : \{0,1\}^n \to \{0,1\}$ be the majority function on the first m bits which is given by $\mathrm{Maj}_m(x^n) := \mathbb{1}\left\{\sum_{i=1}^m x_i \geq m/2\right\}$ for each $x^n \in \{0,1\}^n$. Then, clearly, Maj_1 is a dictator function, and

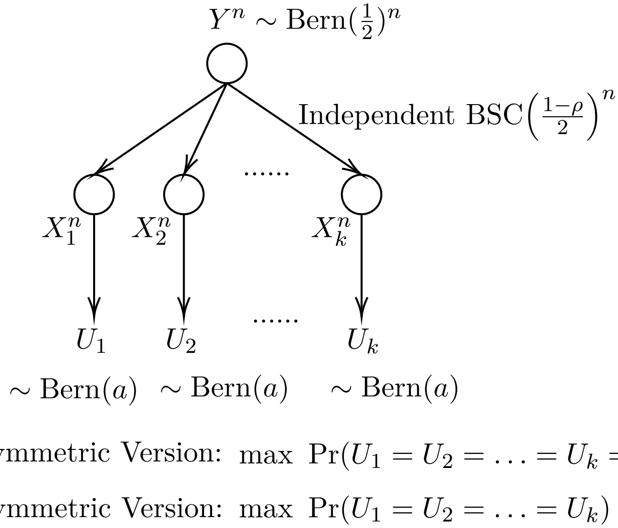

Figure 9.1: The Non-Interactive Correlation Distillation problem with k users

Maj_n is the indicator of the Hamming ball $\mathbb{B}_{n/2}(1^n)$ (as introduced in Section 8.2.2). Hence, majority functions are generalizations of dictator functions and indicators of Hamming balls. Furthermore, we say that a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is *anti-symmetric* (or *odd*) if

$$f(x^n) + f(\bar{x}^n) = 1 \quad \text{for all } x^n \in \{0,1\}^n, \tag{9.1}$$

where $\bar{x}^n := 1^n - x^n$ is the bitwise negation of x^n. Equivalently, for an anti-symmetric Boolean function f, $\mathrm{supp}(f)^{\mathrm{c}} = 1^n - \mathrm{supp}(f)$, where $1^n - \mathcal{A} := \{1^n - x^n : x^n \in \mathcal{A}\}$ for any set $\mathcal{A} \subset \{0,1\}^n$. By definition, for any odd $m \in [n]$, the majority function Maj_m is anti-symmetric.

The k-user NICD problem, which is illustrated in Fig. 9.1, was investigated by Mossel and O'Donnell [122] for the symmetric version, and by Li and Médard [110] for the asymmetric version. There are k correlated memoryless sources $X_1, X_2, \ldots, X_k$ generated from a common memoryless Bernoulli source $Y \sim \mathrm{Bern}(\frac{1}{2})$ through k independent binary symmetric channels with crossover probability $p = (1-\rho)/2$; hence, $0 < \rho < 1$ is the correlation coefficient between $X_{j,i}$ and Y_i for all $j \in [k]$ and $i \in [n]$. A Boolean function $f_i : \{0,1\}^n \to \{0,1\}$ is applied to each source sequence[1] X_i^n to generate a random bit $U_i = f_i(X_i^n)$.

[1]Here, we use the notation X_i^n to denote the i^{th} (out of k) length-n correlated

Definition 9.1. For a dyadic rational $a = M/2^n \in [0,1]$ (in which $M \in \{0,1,\ldots,2^n\}$), define the *forward joint probability at mean* a as

$$\Gamma_\rho^{(k)}(a) := \max_{\substack{\text{Boolean } f_i, 1\le i\le k:\\ \Pr(f_i(X_i^n)=1)=a}} \Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n) = 1\right). \quad (9.2)$$

Since we do not consider the reverse counterpart of the forward joint probability in (9.2) throughout this section, we omit the overline on Γ (cf. the notation $\overline{\Gamma}^{(n)}$ used for the forward joint probability in (8.2)) but we make the number of users k and the correlation coefficient ρ explicit in the notation. To avoid notational overload, we also omit the superscript n that indexes the blocklength. Table 9.1 lists commonly encountered operational quantities in this section.

Table 9.1: Table of commonly used operational quantities in this section

Name	Symbol	Definition(s)
Forward joint probability at a	$\Gamma_\rho^{(k)}(a)$	(9.2), (9.7)
q-stability of f	$\mathbf{S}_\rho^{(q)}[f]$	(9.4)
Asymmetric max q-stability at a	$\Gamma_\rho^{(q)}(a)$	(9.8)
Symmetric max q-stability at a	$\breve{\Gamma}_\rho^{(q)}(a)$	(9.14)
Symmetric q-stability of f	$\breve{\mathbf{S}}_\rho^{(q)}[f]$	(9.15)
Symmetric forward joint probability at a	$\breve{\Gamma}_\rho^{(k)}(a)$	(9.19)
Φ-stability of f	$\mathbf{S}_\rho^{(\Phi)}[f]$	(9.22)
Φ-asymmetric max q-stability at a	$\Pi_\rho^{(q)}(a)$	(9.23)
Φ-symmetric max q-stability at a	$\breve{\Pi}_\rho^{(q)}(a)$	(9.24)
LD exponent	$\Upsilon_{q,\mathrm{LD}}^{(n)}(\alpha)$	(9.51)
MD exponent	$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha)$	(9.52)

It clearly holds that every pair (X_j^n, X_ℓ^n) with $j \neq \ell$ is a source sequence generated by a DSBS with correlation coefficient ρ^2 (because $X_j - Y - X_\ell$). This implies that $\Gamma_\rho^{(2)}(a)$ corresponds to the forward joint probability defined in (8.2) for the DSBS with correlation coefficient ρ^2.

Due to the apparent symmetry of the problem, one may naturally wonder whether the k functions $f_1, \ldots, f_k$ that attain the forward joint

source sequences instead of the random vector $(X_i, X_{i+1}, \ldots, X_n)$.

probability are necessarily identical. This is positively confirmed in the following proposition which can be proved using either the idea in [122, Proposition 3] or [110]. We provide a self-contained proof.

Proposition 9.1. Let $\mathcal{F}$ be any class of Boolean functions. Let $k, n \geq 1$ and $\rho \in (0,1)$. Every tuple of functions $(f_1, \ldots, f_k) \in \mathcal{F}^k$ that maximizes $\Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n) = 1\right)$ satisfies $f_1 = f_2 = \ldots = f_k$.

Proof. Since $\mathcal{F}$ is finite, we may enumerate its elements as $\mathcal{F} = \{g_j : j \in [M]\}$ where $M \geq 2$ to avoid the trivial case in which $M = 1$. Suppose that among the k users, g_j is used by kp_j of them. Then clearly, $\{p_j : j \in [M]\}$ forms a distribution on $\mathcal{F}$ or, isomorphically, on $[M]$. On the other hand, the joint probability induced by this scheme is

$$\Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n) = 1\right) = \mathbb{E}_{Y^n}\left[\prod_{j=1}^{M} (T_\rho g_j(Y^n))^{kp_j}\right], \quad (9.3)$$

where T_ρ is the noise operator defined in (8.55). On the other hand, given $a_1, \ldots, a_M > 0$, the map $(p_1, \ldots, p_M) \in \mathcal{P}([M]) \mapsto \prod_{j=1}^{M} a_j^{p_j}$ is convex. Hence, the expression in (9.3) is *convex* in $(p_1, \ldots, p_M)$. *Maximizing* (9.3) over $(p_1, \ldots, p_M)$ on the probability simplex $\mathcal{P}([M])$, we see that the maximum is attained at a vertex of $\mathcal{P}([M])$. This in turn implies that the maximum of $\Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n) = 1\right)$ over all $(f_1, \ldots, f_k) \in \mathcal{F}^k$ is attained by some $(f_1, \ldots, f_k)$ such that $f_1 = f_2 = \ldots = f_k$. The necessity of the identity of Boolean functions in attaining this maximum can also be verified; see Mossel and O'Donnell [122]. $\square$

By particularizing $\mathcal{F}$ in Proposition 9.2 to be the set of Boolean functions with mean a, any tuple of k functions $(f_1, \ldots, f_k)$ that attains the forward joint probability necessarily satisfies $f_1 = f_2 = \ldots = f_k$. This observation draws our attention to the following related quantity known as the *q-stability* [52], [110].

Definition 9.2. For any $q \in [1, \infty)$ and a Boolean function $f : \{0,1\}^n \to \{0,1\}$, the *q-stability* of f is defined as

$$\mathbf{S}_\rho^{(q)}[f] := \mathbb{E}_{Y^n}\left[(T_\rho f(Y^n))^q\right]. \quad (9.4)$$

For $q = 2$, the q-stability reduces to the *correlation* $\mathbb{E}[f(X^n)f(\hat{X}^n)]$, or equivalently, the *joint probability* $\Pr(f(X^n) = f(\hat{X}^n) = 1)$, where $(X^n, \hat{X}^n)$ is a source sequence of the DSBS with correlation coefficient ρ^2. Hence, $\mathbb{E}[f(X^n)f(\hat{X}^n)]$ is termed the *noise stability* of the Boolean function f with parameter ρ^2, which is denoted as $\mathbf{S}_{\rho^2}[f]$. As mentioned in the discussion following (9.2),

$$\mathbf{S}_{\rho^2}[f] = \mathbf{S}_\rho^{(2)}[f]. \tag{9.5}$$

To better understand the concept of the q-stability, we now compute it for two functions.

Example 9.1. For the dictator function Maj_1,

$$\begin{aligned}\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_1] &= \mathbf{S}_\rho^{(q)}[X_1] \\ &= \mathbb{E}_{Y_1}\left[(\mathbb{E}[X_1|Y_1])^q\right] \\ &= \frac{1}{2}\Big(\frac{1+\rho}{2}\Big)^q + \frac{1}{2}\Big(\frac{1-\rho}{2}\Big)^q.\end{aligned}$$

Example 9.2. For the indicator of the Hamming ball Maj_n, it is not easy to derive the exact value of its q-stability for each dimension $n \in \mathbb{N}$. However, one can determine the limit of the q-stability of Maj_n as $n \to \infty$. By the (multivariate) central limit theorem,

$$\frac{2}{\sqrt{n}}\Bigg(\sum_{i=1}^n \begin{bmatrix} X_i \\ Y_i \end{bmatrix} - \frac{n}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\Bigg) \xrightarrow{\mathrm{d}} \mathcal{N}\Bigg(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{K}\Bigg),$$

where the covariance matrix $\mathbf{K}$ is defined in (8.17). Define the *Gaussian q-stability function* $\Lambda_\rho^{(q)} : [0,1] \to [0,1]$ as

$$\Lambda_\rho^{(q)}(a) := \mathbb{E}\big[\Pr(U \le \Phi^{-1}(a)|V)^q\big] = \mathbb{E}\Bigg[\Phi\Big(\frac{\Phi^{-1}(a) - \rho V}{\sqrt{1-\rho^2}}\Big)^q\Bigg], \tag{9.6}$$

where (U, V) is a pair of jointly Gaussian random variables with zero mean and covariance matrix $\mathbf{K}$. Therefore, for every $(\rho, q) \in (-1, 1) \times [0, 1]$, the limit of the q-stability of Maj_n is

$$\lim_{n\to\infty} \mathbf{S}_\rho^{(q)}[\mathrm{Maj}_n] = \Lambda_\rho^{(q)}(1/2).$$

We relate the q-stability to the NICD problem by observing that for an integer k, the forward joint probability in (9.2) can be rewritten as

$$\Gamma_\rho^{(k)}(a) = \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} \mathbf{S}_\rho^{(k)}[f]. \tag{9.7}$$

Hence, it is natural to term $\Gamma_\rho^{(k)}(a)$ as the *asymmetric max k-stability at mean a*. If we replace the integer k in (9.7) with an arbitrary real number $q \in [1, \infty)$, we can define the asymmetric max q-stability at mean a [52], [110] as follows.

Definition 9.3. For $q \in [1, \infty)$, the *asymmetric max q-stability at mean a* is defined as

$$\Gamma_\rho^{(q)}(a) := \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} \mathbf{S}_\rho^{(q)}[f] \tag{9.8}$$

$$= \max_{\mathcal{A}\subset\{0,1\}^n:\pi_X^n(\mathcal{A})=a} \mathbb{E}_{Y^n}\left[\pi_{X|Y}^n(\mathcal{A}|Y^n)^q\right]. \tag{9.9}$$

The equality in (9.9) follows because for a Boolean function f, $T_\rho f(y^n) = \pi_{X|Y}^n(\mathcal{A}|y^n)$ with $\mathcal{A}$ being the support of f; see (8.55).

A few remarks concerning this definition are in order. First, for fixed $a \in [0,1]$, the function $q \in [1,\infty) \mapsto \Gamma_\rho^{(q)}(a)$ is nonincreasing. Second, given $q \geq 1$ and two correlation coefficients $0 \leq \rho \leq \hat{\rho} \leq 1$, for any $y^n \in \{0,1\}^n$, we have

$$(T_\rho f)^q(y^n) = (T_{\rho/\hat{\rho}} T_{\hat{\rho}} f)^q(y^n) \leq T_{\rho/\hat{\rho}}(T_{\hat{\rho}} f)^q(y^n), \tag{9.10}$$

where the equality follows from the fact that $T_{\rho_1\rho_2} = T_{\rho_1}T_{\rho_2}$ for all $\rho_1, \rho_2 \in [0,1]$, and the inequality follows by Jensen's inequality ($x \mapsto x^q$ is convex for $q \geq 1$). From (9.10), we obtain

$$\mathbf{S}_\rho^{(q)}[f] \leq \mathbb{E}_{Y^n}\left[T_{\rho/\hat{\rho}}(T_{\hat{\rho}} f)^q(Y^n)\right] \tag{9.11}$$

$$= \mathbb{E}_{Z^n}\left[(T_{\hat{\rho}} f)^q(Z^n)\right] \qquad (Z^n \sim \mathrm{Unif}\{0,1\}^n) \tag{9.12}$$

$$= \mathbf{S}_{\hat{\rho}}^{(q)}[f], \tag{9.13}$$

where (9.12) follows because if the input to a binary symmetric channel is uniform, so is its output.[2] Hence, given $q \geq 1$ and $a \in [0,1]$, the function $\rho \in [0,1] \mapsto \Gamma_\rho^{(q)}(a)$ is nondecreasing. Finally, if $\rho = 1$ (i.e., there is no noise), then $\Gamma_1^{(q)}(a) = a$. If instead $\rho = 0$, then $\Gamma_0^{(q)}(a) = a^q$.

[2]The block of inequalities in (9.11)–(9.13) can also be re-interpreted as follows. Given a DSBS (X,Y) with correlation coefficient $\rho \in [0,1]$, we can construct a Markov chain $X - Z - Y$ with correlation coefficient between X and Z being $\hat{\rho} \in [0,\rho]$ such that for any $q \geq 1$, we have $\mathbb{E}[\mathbb{E}[f(X)|Y]^q] \leq \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X)|Z]^q]|Y]] = \mathbb{E}[\mathbb{E}[f(X)|Z]^q]$.

To find the solution to the asymmetric max q-stability problem (i.e., the equivalent optimization problems in Definition 9.3), we have to identify Boolean functions that are the "most stable" under the action of the noise operator T_ρ, with the stability being measured by the q-stability $\mathbf{S}_\rho^{(q)}[f]$.

Analogously to the asymmetric max q-stability at mean a, one can define a *symmetric* version of this stability notion by maximizing the sum of the q-stabilities of f and $1-f$.

Definition 9.4. For $q > 1$, the *symmetric max q-stability at mean a* is[3]

$$\breve{\Gamma}_\rho^{(q)}(a) := \max_{\text{Boolean } f: \mathbb{E}[f(X^n)]=a} \breve{\mathbf{S}}_\rho^{(q)}[f], \tag{9.14}$$

where

$$\breve{\mathbf{S}}_\rho^{(q)}[f] := \mathbf{S}_\rho^{(q)}[f] + \mathbf{S}_\rho^{(q)}[1-f] \tag{9.15}$$

is the *symmetric q-stability of f*.

Let f be an anti-symmetric Boolean function. Consider,

$$\begin{aligned}\mathbf{S}_\rho^{(q)}[1-f] &= \mathbf{S}_\rho^{(q)}[f(1^n - \cdot)] \qquad (9.16)\\ &= \frac{1}{2^n} \sum_{y^n \in \{0,1\}^n} \left(\mathbb{E}[f(\bar{X}^n)|Y^n = y^n]\right)^q \\ &= \frac{1}{2^n} \sum_{\bar{y}^n \in \{0,1\}^n} \left(\mathbb{E}[f(\bar{X}^n)|\bar{Y}^n = \bar{y}^n]\right)^q \\ &= \mathbf{S}_\rho^{(q)}[f], \qquad (9.17)\end{aligned}$$

where (9.16) follows from (9.1), and (9.17) follows because $(\bar{X}^n, \bar{Y}^n)$ has the same joint distribution as (X^n, Y^n). Hence, for an anti-symmetric Boolean function f,

$$\breve{\mathbf{S}}_\rho^{(q)}[f] = 2\,\mathbf{S}_\rho^{(q)}[f]. \tag{9.18}$$

Furthermore, similarly to the asymmetric case, the symmetric max q-stability also admits an important operational interpretation in the

[3] We use the breve accent on symbols that signify *symmetric* quantities (e.g., $\breve{\Gamma}_\rho^{(q)}$). The breve serves as a mnemonic as it is symmetric about a vertical axis.

k-user NICD problem; see (9.19). To describe this, we need to first introduce the following proposition, which is the symmetric counterpart of Proposition 9.1 and is due to Mossel and O'Donnell [122]. The proof is almost the same as that of Proposition 9.1 and hence, is omitted.

Proposition 9.2. Let $\mathcal{F}$ be any class of Boolean functions. Let $k, n \geq 1$ and $\rho \in (0, 1)$. Every tuple of functions $(f_1, \ldots, f_k) \in \mathcal{F}^k$ that maximizes $\Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n)\right)$ satisfies $f_1 = f_2 = \ldots = f_k$.

By choosing $\mathcal{F}$ in Proposition 9.2 to be the set of Boolean functions with mean a, we deduce that the symmetric max q-stability with $q = k$ (an integer) satisfies

$$\breve{\Gamma}_\rho^{(k)}(a) = \max_{\substack{\text{Boolean } f_i, 1 \leq i \leq k: \\ \Pr(f_i(X_i^n)=1)=a}} \Pr\left(f_1(X_1^n) = \ldots = f_k(X_k^n)\right). \quad (9.19)$$

This is also called the *symmetric forward joint probability* in the k-user NICD problem. Thus, the symmetric max k-stability $\breve{\Gamma}_\rho^{(k)}$ quantifies the *maximum agreement probability* over all Boolean functions with a fixed mean in the k-user NICD problem (Fig. 9.1). In contrast, the forward joint probability or asymmetric max k-stability $\Gamma_\rho^{(k)}$ (in (9.2) and (9.7)) quantifies the maximum agreement probability *when the generated bits take on the value* 1.

9.1.2 Variants of q-Stabilities

The reader will notice that the definitions of the asymmetric and symmetric max q-stabilities in (9.9) and (9.14) are trivial for the case $q = 1$, since for this case, any Boolean f such that $\mathbb{E}[f(X^n)] = a$ satisfies

$$\mathbf{S}_\rho^{(1)}[f] = a \quad \text{and} \quad \breve{\mathbf{S}}_\rho^{(1)}[f] = 1. \quad (9.20)$$

Hence, the asymmetric and symmetric max 1-stabilities at mean a are attained by *any* Boolean functions with mean a. Are there any "more meaningful" notions of asymmetric and symmetric max q-stabilities for $q = 1$? We answer this question in the affirmative by defining variants of the max q-stabilities. These variants connect the q-stabilities to the *most informative Boolean functions* problem of Courtade and Kumar

[40], one of the most important open problems in information theory at the time of the writing of this monograph.

To introduce these variants, for $q \geq 1$, define[4] $\Phi_q, \breve{\Phi}_q : (0,1) \to \mathbb{R}$ as

$$\Phi_q(t) := t \cdot \frac{\ln_q(t)}{\ln 2} \quad \text{and} \quad \breve{\Phi}_q(t) := \Phi_q(t) + \Phi_q(1-t), \tag{9.21}$$

where $\ln_q : (0,\infty) \to \mathbb{R}$ is defined as

$$\ln_q(t) := \begin{cases} \ln(t) & q = 1 \\ \dfrac{t^{q-1}-1}{q-1} & q > 1 \end{cases}$$

and is known as the *q-logarithm* introduced by Tsallis [164], but with a slight reparameterization. Note that for Φ_q and $\breve{\Phi}_q$, the case of $q = 1$ is the continuous extension of the case $q > 1$.

Definition 9.5. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and another function $\Phi : (0,1) \to \mathbb{R}$, define the *$\Phi$-stability of f* with respect to a correlation parameter ρ as

$$\mathbf{S}_\rho^{(\Phi)}[f] = \mathbb{E}_{Y^n}\left[\Phi(T_\rho f(Y^n))\right]. \tag{9.22}$$

Thus, this definition is analogous to that of the q-stability (Definition 9.2) as we recover the latter when we instantiate $\Phi(t) = t^q$. We are, however, going to consider Φ to be the functions in (9.21).

Definition 9.6. Define the *Φ-asymmetric* and *Φ-symmetric max q-stabilities at mean a* as

$$\Pi_\rho^{(q)}(a) := \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} \mathbf{S}_\rho^{(\Phi_q)}[f] \quad \text{and} \tag{9.23}$$

$$\breve{\Pi}_\rho^{(q)}(a) := \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} \mathbf{S}_\rho^{(\breve{\Phi}_q)}[f]. \tag{9.24}$$

For $q > 1$, it is easy to verify that

$$\Pi_\rho^{(q)}(a) = \frac{\Gamma_\rho^{(q)}(a) - a}{(q-1)\ln 2} \quad \text{and} \quad \breve{\Pi}_\rho^{(q)}(a) = \frac{\breve{\Gamma}_\rho^{(q)}(a) - 1}{(q-1)\ln 2}, \tag{9.25}$$

[4] These functions are not to be confused with the Gaussian cumulative distribution function which is also denoted as $\Phi(\cdot)$.

where $\Gamma_\rho^{(q)}(a)$ and $\breve{\Gamma}_\rho^{(q)}$ are the asymmetric and symmetric max q-stabilities defined in (9.9) and (9.14) respectively. For $q = 1$, the Φ-asymmetric and Φ-symmetric max 1-stabilities at mean a can be expressed respectively as

$$\Pi_\rho^{(1)}(a) = \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} \mathbb{E}_{Y^n}\left[T_\rho f(Y^n) \log T_\rho f(Y^n)\right], \tag{9.26}$$

and

$$\breve{\Pi}_\rho^{(1)}(a) = \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} -H(f(X^n)|Y^n) \tag{9.27}$$

$$= \max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=a} I(f(X^n);Y^n) - h(a). \tag{9.28}$$

The objective function in (9.26) is known as the *entropy (functional)* of the noisy Boolean function $T_\rho f$, and the objective function in (9.27) is the negative *conditional Shannon entropy* of $f(X^n)$ given Y^n. For dictator functions f, the conditional Shannon entropy $H(f(X^n)|Y^n)$ is equal to $H(X_1|Y_1) = h((1-\rho)/2)$.

The maximization in (9.28) for $a = 1/2$ corresponds to the balanced version of the *most informative Boolean function* problem which was first studied in the papers of Courtade and Kumar [40], [102]. They conjectured that the maximum in (9.28) is attained by dictator functions (cf. Section 8.2.1). In the following section, we provide more details on this conjecture, and we also review several related conjectures concerning the q-stabilities. Observe from the one-to-one relationships in (9.25) that for $q > 1$, the original definitions of the asymmetric and symmetric max q-stabilities $\Gamma_\rho^{(q)}$ and $\breve{\Gamma}_\rho^{(q)}$ in (9.9) and (9.14) are "equivalent" to their Φ-versions $\Pi_\rho^{(q)}$ and $\breve{\Pi}_\rho^{(q)}$ defined respectively in (9.23) and (9.24), in the sense that once the former (resp. the latter) has been determined, the latter (resp. the former) will also be determined. Hence, throughout this section, for $q > 1$, we refer to $\Gamma_\rho^{(q)}$ and $\Pi_\rho^{(q)}$ interchangeably for the asymmetric case. We will also refer to $\breve{\Gamma}_\rho^{(q)}$ and $\breve{\Pi}_\rho^{(q)}$ interchangeably for the symmetric case. However, for $q = 1$, we only consider the quantities $\Pi_\rho^{(1)}$ and $\breve{\Pi}_\rho^{(1)}$ since the definitions of $\Gamma_\rho^{(1)}$ and $\breve{\Gamma}_\rho^{(1)}$ are trivial for this case; see (9.20).

9.2 Related Conjectures

In this section, we introduce several prominent conjectures on the max q-stabilities. We first consider the optimality of dictator functions in attaining the asymmetric and symmetric max q-stabilities at mean $a = 1/2$ (also called the *balanced* case). For ease of reference, we first state a corollary to Witsenhausen's classical result [178] in Theorem 8.3. This corollary implies that dictator functions are optimal in attaining asymmetric or symmetric max q-stabilities for $q = 2$ and $a = 1/2$ (the 2-user NICD problem in the CL regime with $a = b = 1/2$).

Corollary 9.1. For $q = 2$ and $\rho \in (0, 1)$, both $\Gamma_\rho^{(q)}(1/2)$ and $\breve{\Gamma}_\rho^{(q)}(1/2)$ are attained by dictator functions.

There was no further progress on the max q-stability problem for almost 30 years since Witsenhausen's seminal work [178] in 1975. In 2005, Mossel and O'Donnell [122] considered the symmetric max q-stability problem with $q \in \{3, 4, 5, \ldots\}$ and made progress on this problem. They resolved the case of $q = 3$ for the balanced case (i.e., $a = 1/2$) using a cute reduction argument.

Theorem 9.2. For $q = 3$ and $\rho \in (0, 1)$, $\breve{\Gamma}_\rho^{(q)}(1/2)$ is attained by dictator functions.

Proof. Theorem 9.2 can be proved by reducing the problem involving $q = 3$ to the (simpler) problem in which $q = 2$. By the equivalence between the NICD problem and max q-stability, we consider the 3-user NICD problem with 3 (possibly) distinct functions (f_1, f_2, f_3). For brevity, denote the values of the joint probability mass function of $(U_1, U_2, U_3) = (f_1(X_1^n), f_2(X_2^n), f_3(X_3^n))$ as $\{p_{000}, p_{001}, \ldots, p_{111}\}$. Then, we see that the following identity holds:

$$3 + \sum_{(i,j)\in[3]^2 : i\neq j} \Pr(U_i = U_j) = 5 + 4 \Pr(U_1 = U_2 = U_3). \quad (9.29)$$

This identity can be verified by bookkeeping the probability masses. For example, note that $\Pr(U_1 = U_2) = p_{000} + p_{001} + p_{110} + p_{111}$ and $\Pr(U_1 = U_2 = U_3) = p_{000} + p_{111}$. Having established this, leveraging the case for $q = 2$ (Corollary 9.1), we know that the left-hand side of (9.29) is maximized by identical dictator functions over all balanced Boolean functions (i.e., $\mathbb{E}[f_i(X_i^n)] = 1/2$); hence, so is the right-hand side. □

Based on Corollary 9.1 and Theorem 9.2, one may naïvely conjecture that dictator functions are optimal in attaining the asymmetric or symmetric max q-stability at mean $1/2$ for any integer $q \geq 2$. However, this was disproved by Mossel and O'Donnell [122]. Specifically, using computer-assisted calculations, they constructed counterexamples, as shown in the following proposition, such that when $q = 10$, dictator functions are not optimal in attaining the asymmetric and symmetric max q-stabilities at mean $a = 1/2$.

Proposition 9.3. For $q = 10$ and $\rho = 0.48$, it holds that

$$\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_3] > \max\left\{\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_1], \mathbf{S}_\rho^{(q)}[\mathrm{Maj}_5]\right\} \quad \text{and} \tag{9.30}$$
$$\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_3] > \max\left\{\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_1], \breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_5]\right\}. \tag{9.31}$$

Proof. By computer-assisted calculations, for $q = 10$ and $\rho = 0.48$, one finds that $\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_1] \leq 0.0493$, $\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_5] \leq 0.0488$, and $\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_3] \geq 0.0496$. Hence, the inequality in (9.31) holds. The inequality in (9.30) follows from (9.31) since $\breve{\mathbf{S}}_\rho^{(q)}[\mathrm{Maj}_m] = 2\,\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_m]$ for any odd $m \leq n$; see (9.18). □

Since $\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_3] > \mathbf{S}_\rho^{(q)}[\mathrm{Maj}_1]$ and Maj_1 is a dictator function, dictators are not optimal for $q = 10$ and $\rho = 0.48$. Furthermore, since the indicators of subcubes and the indicators of Hamming balls (or spheres) have been shown to be optimal or asymptotically optimal in several cases for the NICD problem (Sections 8.3–8.5), one may wonder whether the max q-stability is always exactly attained by these functions. The inequality $\mathbf{S}_\rho^{(q)}[\mathrm{Maj}_3] > \mathbf{S}_\rho^{(q)}[\mathrm{Maj}_5]$ implies a negative answer to this question. For $n = 5$, $q = 10$, $a = 1/2$, and $\rho = 0.48$, both the indicators of subcubes and Hamming balls in the 5-dimensional Hamming cube are not optimal. In fact, Maj_3 corresponds to the indicator of a set formed by multiplying Hamming balls in the 3-dimensional cube and the 2-dimensional cube.

Now things have become relatively clearer. For small q, e.g., $q = 2$ or $q = 3$, dictator functions are optimal in attaining the asymmetric (for $q = 2$) or symmetric (for $q = 2, 3$) max q-stabilities at mean $a = 1/2$. On the other hand, for large q, e.g., $q = 10$, dictator functions are not optimal. Mossel and O'Donnell [122] conjectured that for all $q \in \{4, 5, \ldots, 9\}$, dictator functions maximize the symmetric q-stability $\breve{\Gamma}_\rho^{(q)}(1/2)$ over all balanced Boolean functions.

The symmetric max 1-stability problem at mean $a = 1/2$ (i.e., $\breve{\Pi}_\rho^{(1)}(1/2)$) was studied by Kumar and Courtade [102] and [40]. This question concerns the identification of the class of balanced Boolean functions that maximize the mutual information $I(f(Y^n); X^n)$; cf. (9.28). The authors conjectured that dictator functions maximize the symmetric 1-stability. We note that this is a weaker version of the original conjecture posed by Courtade and Kumar. In the original version of their conjecture, the Boolean functions are not restricted to be balanced. Along these lines, Li and Médard [110] conjectured that for $q \in (1, 2)$ (non-integer), the max asymmetric q-stability is still attained by dictator functions. Here we summarize and generalize this family of conjectures in the following two conjectures.

Conjecture 9.1 (Asymmetric max q-stability). For $\rho \in [0, 1]$ and $q \in [1, 9]$, $\Pi_\rho^{(q)}(1/2)$ is attained by dictator functions.

Conjecture 9.2 (Symmetric max q-stability). For $\rho \in [0, 1]$ and $q \in [1, 9]$, $\breve{\Pi}_\rho^{(q)}(1/2)$ is attained by dictator functions.

Observe that dictator functions are anti-symmetric. Hence, (9.18) holds for dictator functions, which implies that if Conjecture 9.1 is true, so is Conjecture 9.2. Conjectures 9.1 and 9.2 together consist of three (named) conjectures, as summarized in Table 9.2.

Barnes and Özgür [11] proved an interesting dichotomy concerning these conjectures.

Lemma 9.3. For $a = 1/2$, there are two thresholds $q_{\min}$ and $q_{\max}$ satisfying $1 \le q_{\min} \le 2 \le q_{\max}$ such that dictator functions are optimal in attaining the asymmetric max q-stability with $q \ge 1$ if and only if $q \in [q_{\min}, q_{\max}]$. This statement also holds for the symmetric max q-stability but with possibly different thresholds $\breve{q}_{\min}$ and $\breve{q}_{\max}$ satisfying the same condition $1 \le \breve{q}_{\min} \le 2 \le \breve{q}_{\max}$.

Table 9.2: Illustration of the various named conjectures on max q-stabilities; these constitute Conjectures 9.1 and 9.2

q	Are dictators optimal in attaining $\Pi_\rho^{(q)}(1/2)$ (or $\breve{\Pi}_\rho^{(q)}(1/2)$)?
$q = 1$	Courtade–Kumar conjecture (balanced version) [40]
$1 < q < 2$	Li–Médard conjecture [110]
$q = 2$ (2-User NICD)	True and shown by Witsenhausen [178] (cf. Section 8.3.1)
$2 < q \leq 9$	Mossel–O'Donnell conjecture [122]

Proof Sketch of Lemma 9.3. For any $q \in \mathbb{R}$ (not necessarily greater than or equal to 1), define

$$N_q(f) := 2^n \|T_\rho f\|_q^q = \sum_{y^n \in \{0,1\}^n} (T_\rho f(y^n))^q.$$

Let f_0 be a dictator function, e.g., $f_0 = \mathrm{Maj}_1$. Define

$$g_f(q) := N_q(f) - N_q(f_0). \tag{9.32}$$

By using a result due to Laguerre [104], one can find that the sum of exponentials $g_f(q)$ has at most four roots. Observe that

$$g_f(0) = 0, \quad g_f(1) = 0, \quad g_f(2) \leq 0, \quad \text{and}$$
$$g_f(q),\ g_f(-q) > 0 \quad \text{for sufficiently large } q.$$

From these observations, we know that $g_f(q)$ has a root at $q_1 \geq 2$, another at $q_2 = 1$, and another at $q_3 = 0$. Moreover, the remaining root q_4 satisfies $q_4 \leq 2$. Hence, $g_f(q) \leq 0$ for all q in the interval $[\max\{q_4, 1\}, q_1]$; see Fig. 9.2. Taking the intersection of these intervals for all non-dictator functions f, we obtain the interval $[q_{\min}, q_{\max}]$, where $q_{\min}$ and $q_{\max}$ are the desired thresholds. The symmetric case follows similarly. $\square$

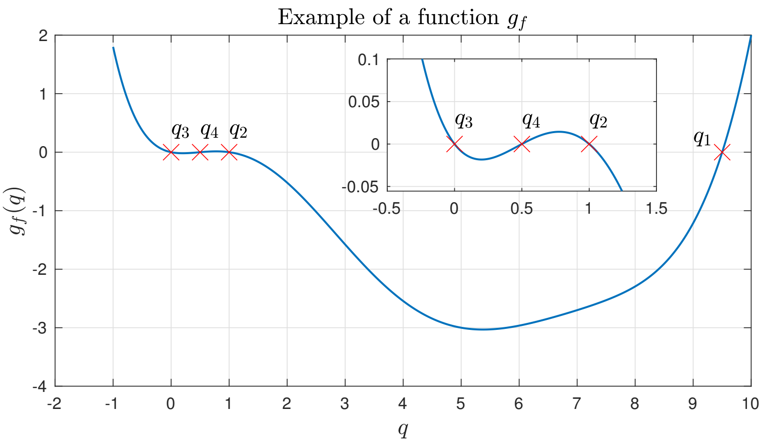

Figure 9.2: Example of a function g_f in (9.32).

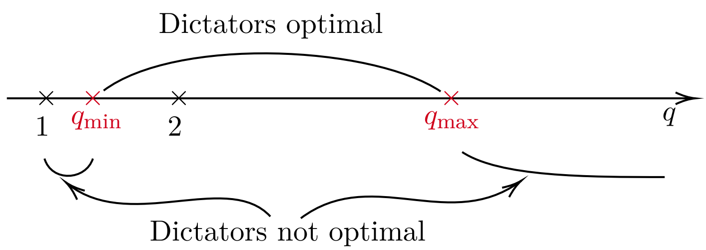

Figure 9.3: Illustration of Lemma 9.3.

This lemma is illustrated in Fig. 9.3.

Remark 9.1. This lemma has several important implications.

(a) Firstly, this lemma implies that both the Courtade–Kumar conjecture and the Li–Médard conjecture are equivalent to the statements that $q_{\min} = 1$ for the asymmetric version and $\breve{q}_{\min} = 1$ for the symmetric version. Hence, the Courtade–Kumar conjecture and the Li–Médard conjecture are also equivalent (to each other).

(b) Secondly, it also implies that the Mossel–O'Donnell conjecture is equivalent to the statements that $q_{\max} \geq 9$ for the asymmetric version and $\breve{q}_{\max} \geq 9$ for the symmetric version. On the other hand, from Proposition 9.3, we see that $\max\{q_{\max}, \breve{q}_{\max}\} < 10$.

(c) Lastly, by Lemma 9.3 and Theorem 9.2 (i.e., Conjecture 9.2 holds for $q = 3$), Conjecture 9.2 also holds for any $q \in [2, 3]$. In other words, Conjecture 9.2 is only open for $q \in [1, 2) \cup (3, 9]$.

Combining all points in Remark 9.1 yields that $2 \le q_{\max} < 10$ and $3 \le \breve{q}_{\max} < 10$. If Conjectures 9.1 and 9.2 hold, then the estimates of $q_{\max}$ and $\breve{q}_{\max}$ can be improved to $9 \le q_{\max}, \breve{q}_{\max} < 10$, as shown in Fig. 9.4.

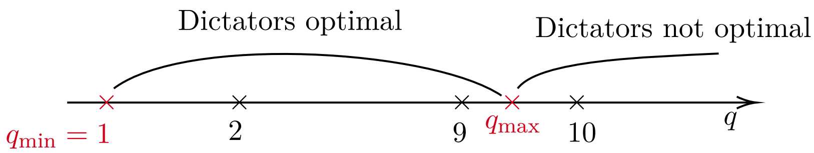

Figure 9.4: Illustration of the range of q for the optimality of dictator functions if Conjectures 9.1 and 9.2 are true.

9.3 Extreme Cases of the Correlation Coefficient

To better understand the max q-stabilities, and also to connect them to several well-known concepts in the analysis of Boolean functions, we first focus our attention on the extreme cases in which the correlation coefficient ρ tends to 0 or 1, but the dimension (or blocklength) n is kept fixed. To illustrate the intuition as to why some results hold, we introduce the concepts of *influences* and *edge-isoperimetric inequalities*.

9.3.1 Influences

For a vector $x^n \in \{0, 1\}^n$, we denote the vector with the i^{th} bit flipped as $(x^n)^{\oplus i} := (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$. Denote the length-$(n - 1)$ with the i^{th} component removed as $x^{\setminus i} := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n)$.

Definition 9.7. The *influence of coordinate* $i \in [n]$ on a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ is defined as

$$\mathbf{I}_i[f] := \Pr\left(f(X^n) \neq f((X^n)^{\oplus i})\right),$$

where $X^n \sim \text{Unif}\{0, 1\}^n$.

Let f be a Boolean function that depends only on $x^{\setminus i}$. This means that the value of f evaluated at every x^n is independent of the i^{th} component x_i. For such an f, clearly, $\mathbf{I}_i[f] = 0$. On the other hand, if f depends only on the i^{th} component, i.e., the dictator functions $f(x^n) = x_i$ or $1 - x_i$, then $\mathbf{I}_i[f] = 1$. Hence, the influence of coordinate i measures how much a function is influenced by the i^{th} coordinate of the input; this coincides with the literal meaning of "influence". Furthermore, the influence can be also expressed in terms of the discrete derivative operator as follows:

Definition 9.8. Let $x^{i\mapsto b} := (x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$. The i^{th} *discrete derivative operator* D_i maps a function $f : \{0,1\}^n \to \mathbb{R}$ to the function $D_i f : \{0,1\}^n \to \mathbb{R}$ defined as

$$D_i f(x^n) := f(x^{i\mapsto 1}) - f(x^{i\mapsto 0}).$$

One observes that

$$\mathbf{I}_i[f] = \mathbb{E}\big[D_i f(X^n)^2\big] = \|D_i f\|_2^2. \tag{9.33}$$

This formula enables us to generalize the definition of the influence from a Boolean function to an arbitrary real-valued function defined on $\{0,1\}^n$; see O'Donnell [131]. We do not discuss this generalization here, since we only mainly focus on Boolean functions.

Definition 9.9. The *total influence* (or *average sensitivity*) of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined as

$$\mathbf{I}[f] := \sum_{i=1}^{n} \mathbf{I}_i[f].$$

The quantities $D_i f$, $\mathbf{I}_i[f]$, and $\mathbf{I}[f]$ admit the following Fourier-analytic representations.

Theorem 9.4. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and $i \in [n]$,

$$D_i f(x^n) = -2 \sum_{\mathcal{S}\subset[n]:\mathcal{S}\ni i} \hat{f}_{\mathcal{S}} \cdot \chi_{\mathcal{S}\setminus\{i\}}(x^n), \tag{9.34}$$

$$\mathbf{I}_i[f] = 4 \sum_{\mathcal{S}\subset[n]:\mathcal{S}\ni i} \hat{f}_{\mathcal{S}}^2, \qquad \text{and} \tag{9.35}$$

$$\mathbf{I}[f] = 4 \sum_{\mathcal{S}\subset[n]} |\mathcal{S}| \hat{f}_{\mathcal{S}}^2 = 4 \sum_{k=0}^{n} k \cdot \mathbf{W}_k[f]. \tag{9.36}$$

where $\mathbf{W}_k[f]$ denotes the degree-k Fourier weight of f defined in (8.60).

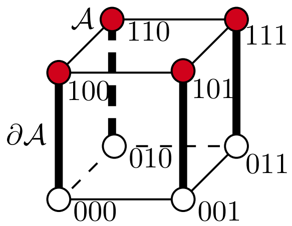

Figure 9.5: Hamming graph for $n = 3$. For the dictator function $\mathrm{Maj}_1(x^3) = x_1$, the set $\mathcal{A} = \{(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$ (indicated in red balls). The edge boundary $\partial\mathcal{A}$ of $\mathcal{A}$ is indicated as the four thick edges. Each boundary edge is a dimension-1 edge.

Proof. Since D_i is a linear operator, (9.34) follows by expressing f in terms of its Fourier coefficients, and then applying the following identity

$$D_i \chi_{\mathcal{S}}(x^n) = \begin{cases} -2\,\chi_{\mathcal{S}\setminus\{i\}}(x^n) & i \in \mathcal{S} \\ 0 & i \notin \mathcal{S} \end{cases}.$$

The identity in (9.35) follows from (9.33) and (9.34), and the fact that for any sets $\mathcal{S}, \mathcal{T} \subset [n]$,

$$\mathbb{E}[\chi_{\mathcal{S}}(X^n)\chi_{\mathcal{T}}(X^n)] = \mathbb{1}\{\mathcal{S} = \mathcal{T}\}. \tag{9.37}$$

The identity in (9.36) follows from (9.35) and Definition 9.9. □

The quantities $\mathbf{I}_i[f]$ and $\mathbf{I}[f]$ also admit interesting graph-theoretic interpretations. Consider the undirected graph in which the vertices consist of all vectors in $\{0,1\}^n$, and two vertices $x^n, y^n \in \{0,1\}^n$ are joined by an edge if the Hamming distance between them is exactly 1, i.e., $d_{\mathrm{H}}(x^n, y^n) = 1$. This graph is known as the *Hamming graph*; see Fig. 9.5 for the Hamming graph when $n = 3$.

Definition 9.10. For a set $\mathcal{A} \subset \{0,1\}^n$, define its *edge boundary* $\partial\mathcal{A}$ as the set of edges in the Hamming graph such that one of its endpoints belongs to $\mathcal{A}$ while the other one belongs to $\mathcal{A}^{\mathrm{c}}$. Every edge that belongs to $\partial\mathcal{A}$ is called a *boundary edge*. An boundary edge $\{x^n, y^n\} \in \partial\mathcal{A}$ is known as a *dimension-i edge* if $y^n = (x^n)^{\oplus i}$, i.e., x^n and y^n are identical except in their i^{th} coordinates.

For a set $\mathcal{A} \subset \{0,1\}^n$, one observes the following facts.

1. The fraction of dimension-i edges that are boundary edges of $\mathcal{A}$ in the Hamming graph is equal to $\mathbf{I}_i[\mathbb{1}_{\mathcal{A}}]$.

2. The fraction of edges in the Hamming graph that are boundary edges of $\mathcal{A}$ is equal to $\frac{1}{n}\mathbf{I}[\mathbb{1}_{\mathcal{A}}]$. This implies that $|\partial\mathcal{A}| = 2^{n-1}\mathbf{I}[\mathbb{1}_{\mathcal{A}}]$, since the total number of edges in the Hamming graph is $n\,2^{n-1}$.

Example 9.3. Let $n = 3$. The Hamming graph is shown in Fig. 9.5. This graph has $3\cdot 2^{3-1} = 12$ edges. Consider the dictator function $\mathrm{Maj}_1(x^3) = x_1$. The support of f is the set $\mathcal{A} = \{(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$. Both $\mathcal{A}$ and $\partial\mathcal{A}$ are indicated in Fig. 9.5 and $|\partial\mathcal{A}| = 4$. For this dictator function, $\mathbf{I}_1[\mathbb{1}_{\mathcal{A}}] = 1$ and $\mathbf{I}_2[\mathbb{1}_{\mathcal{A}}] = \mathbf{I}_3[\mathbb{1}_{\mathcal{A}}] = 0$ (as discussed after Definition 9.7). Note from Fig. 9.5 that there are four dimension-1 edges and no dimension-2 and dimension-3 edges. Thus, the *fractions* of dimension-1, dimension-2 and dimension-3 edges that are boundary edges of $\mathcal{A}$ are 1, 0, and 0 respectively, corroborating Fact 1. Furthermore, $\mathbf{I}[\mathbb{1}_{\mathcal{A}}] = \sum_{i=1}^{3} \mathbf{I}_i[\mathbb{1}_{\mathcal{A}}] = 1$ and the fraction of edges that belong to $\partial\mathcal{A}$ is $1/3 = 4/12$, corroborating Fact 2.

9.3.2 Edge-Isoperimetric Inequalities

From Fact 2, we see that the total influence of f is related to the cardinality of the edge boundary of its support set $\mathcal{A}$. A classical result due to Harper [74] quantifies this relation via the so-called *edge-isoperimetric inequality*.

Theorem 9.5 (Edge-isoperimetric inequality). For $f : \{0,1\}^n \to \{0,1\}$ with $a = \min\{\mathbb{E}[f], 1-\mathbb{E}[f]\}$,

$$\mathbf{I}[f] \geq 2a \log\left(\frac{1}{a}\right). \tag{9.38}$$

This inequality can be seen as a Boolean function version of the *log-Sobolev inequality*. The relationship between this edge-isoperimetric inequality and the real-valued function version of log-Sobolev inequalities will be discussed extensively in Section 10.4.

This inequality in (9.38) is sharp for $a = 2^{-k}$ with $1 \leq k \leq n$, since for this case, the indicator function of an $(n-k)$-subcube attains the lower bound. If a, a dyadic rational, is the mean of f, $\mathbf{I}[f]$ is minimized when f is the indicator of a lexicographic set of size $2^n a$ (cf. Section 8.2.1). The edge-isoperimetric inequality will be used to resolve the extreme cases of the max q-stability problem via the following two theorems that establish a connection between the q-stability and the total influence.

Theorem 9.6. For $f : \{0,1\}^n \to \{0,1\}$,

$$\mathbf{S}_\rho^{(2)}[f] = \sum_{\mathcal{S} \subset [n]} \rho^{2|\mathcal{S}|} \, \hat{f}_{\mathcal{S}}^2 = \sum_{k=0}^{n} \rho^{2k} \, \mathbf{W}_k[f].$$

Proof. This theorem follows by (9.37) and the facts that

$$\mathbf{S}_\rho^{(2)}[f] = \langle T_\rho f, T_\rho f \rangle \qquad \text{and} \qquad (\widehat{T_\rho f})_{\mathcal{S}} = \sum_{\mathcal{S} \subset [n]} \rho^{|\mathcal{S}|} \, \hat{f}_{\mathcal{S}}, \tag{9.39}$$

where $\{(\widehat{T_\rho f})_{\mathcal{S}}\}_{\mathcal{S} \subset [n]}$ are the Fourier coefficients of $T_\rho f$. ∎

This theorem implies that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\rho} \mathbf{S}_\rho^{(2)}[f] \Big|_{\rho=1} &= \frac{\mathbf{I}[f]}{2}, \\ \frac{\mathrm{d}}{\mathrm{d}\rho} \mathbf{S}_\rho^{(2)}[f] \Big|_{\rho=0} &= 0, \qquad \text{and} \\ \frac{\mathrm{d}^2}{\mathrm{d}\rho^2} \mathbf{S}_\rho^{(2)}[f] \Big|_{\rho=0} &= 2\, \mathbf{W}_1[f]. \end{aligned}$$

Theorem 9.6 pertains to $q = 2$. For general $q > 1$, the derivatives of $\mathbf{S}_\rho^{(q)}[f]$ at $\rho = 0$ and 1 are given in the following theorem which is due to Li and Médard [110].

Theorem 9.7. For $f : \{0,1\}^n \to \{0,1\}$ with mean $a \in (0,1]$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\rho} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=1} &= \frac{q}{4} \, \mathbf{I}[f], \qquad (9.40) \\ \frac{\mathrm{d}}{\mathrm{d}\rho} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=0} &= 0, \qquad \text{and} \\ \frac{\mathrm{d}^2}{\mathrm{d}\rho^2} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=0} &= q(q-1) \, a^{q-2} \, \mathbf{W}_1[f]. \end{aligned}$$

Proof. By using the Fourier-analytic relations in (9.39), we obtain

$$T_\rho f(y^n) = \sum_{\mathcal{S}\subset[n]} \rho^{|\mathcal{S}|}\, \hat{f}_{\mathcal{S}}\, \chi_{\mathcal{S}}(y^n).$$

By the definition of the q-stability in (9.4)

$$\mathbf{S}_\rho^{(q)}[f] = \mathbb{E}_{Y^n}\left[\left(\sum_{\mathcal{S}\subset[n]} \rho^{|\mathcal{S}|}\hat{f}_{\mathcal{S}}\chi_{\mathcal{S}}(Y^n)\right)^q\right].$$

Differentiating this with respect to ρ yields

$$\frac{\mathrm{d}}{\mathrm{d}\rho}\mathbf{S}_\rho^{(q)}[f] = q\ \mathbb{E}_{Y^n}\left[\left(T_\rho f(Y^n)\right)^{q-1} \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|\geq 1} |\mathcal{S}|\rho^{|\mathcal{S}|-1}\hat{f}_{\mathcal{S}}\chi_{\mathcal{S}}(Y^n)\right].$$

Setting $\rho = 1$, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\rho}\mathbf{S}_\rho^{(q)}[f]\Big|_{\rho=1} &= q\ \mathbb{E}_{Y^n}\left[f(Y^n)^{q-1} \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|\geq 1} |\mathcal{S}|\hat{f}_{\mathcal{S}}\chi_{\mathcal{S}}(Y^n)\right] \\
&= q\ \mathbb{E}_{Y^n}\left[f(Y^n) \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|\geq 1} |\mathcal{S}|\hat{f}_{\mathcal{S}}\chi_{\mathcal{S}}(Y^n)\right] \qquad (9.41) \\
&= q \sum_{\mathcal{S}\subset[n]:|\mathcal{S}|\geq 1} |\mathcal{S}|\hat{f}_{\mathcal{S}}^2 \\
&= \frac{q}{4}\ \mathbf{I}[f], \qquad (9.42)
\end{aligned}$$

where (9.41) follows since f only takes values in $\{0,1\}$, and hence, $f^{q-1} = f$, and (9.42) follows from (9.36). This proves (9.40). The other equalities can be proved similarly. □

9.3.3 Max q-Stabilities in Extreme Cases of ρ

Based on the concept of the total influence and the results stated in Sections 9.3.1 and 9.3.2, we are now ready to analyze the extreme cases of the max q-stability as $\rho \downarrow 0$ and $\rho \uparrow 1$. We first state a lower bound on the derivative of the q-stability with respect to ρ evaluated at $\rho = 1$. This result is due to Mossel and O'Donnell [122] for integer q and Li and Médard [110] for real $q > 1$.

Theorem 9.8. Let $q > 1$. For a Boolean function $f : \{0,1\}^n \to \{0,1\}$ with mean a,

$$\frac{\partial}{\partial \rho} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=1} \geq \frac{q}{2}\, a \,\log\left(\frac{1}{a}\right).$$

This lower bound is attained if $a = 2^{-k}$ for any $1 \leq k \leq n$ and f is the indicator of an $(n-k)$-subcube.

This theorem follows by the edge-isoperimetric inequality in (9.38) and (9.40).

Note that if $\rho = 1$, then for any $f : \{0,1\}^n \to \{0,1\}$ with mean a, it holds that $\mathbf{S}_\rho^{(q)}[f] = a$ (cf. (9.20)). Hence, from Theorem 9.8, it is plausible, via "continuity arguments", that if $a = 2^{-k}$ for integer k and ρ is sufficiently close to 1, then $\mathbf{S}_\rho^{(q)}[f]$ is maximized by the indicator of an $(n-k)$-subcube. This can be proven rigorously using the fact that the number of Boolean functions for a given n is finite, some approximation arguments involving Taylor's theorem, and bounds on the derivative of $\mathbf{S}_\rho^{(q)}$ evaluated at $\rho = 1$ (Theorem 9.8). This is stated formally in the following theorem which is due to Mossel and O'Donnell [122] for integer q and Li and Médard [110] for real q.

Theorem 9.9. Fix $n \geq 1$, $q > 1$, and $a = 2^{-k}$ with $1 \leq k \leq n$. There exists an $\epsilon \in (0,1)$ such that for all $\rho \in [1-\epsilon, 1]$, $\Gamma_\rho^{(q)}(a)$ is attained by the indicator of an $(n-k)$-subcube.

Proof Sketch of Theorem 9.9. Fix a Boolean function f and $\rho \in (0,1)$. Using Taylor's theorem, we can write

$$\mathbf{S}_\rho^{(q)}[f] = \mathbf{S}_1^{(q)}[f] + (\rho - 1)\frac{\partial}{\partial \rho}\mathbf{S}_\rho^{(q)}[f]\Big|_{\rho=1} + \phi_f(\tilde{\rho})(\rho-1)^2,$$

where $\phi_f : [0,1] \to \mathbb{R}$ is a bounded function induced by f and $\tilde{\rho} \in (\rho, 1)$. Since n is fixed, the number of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ is finite. From this fact, we deduce that $\phi(\rho) := \max_{f:\{0,1\}^n \to \{0,1\}} \phi_f(\rho)$, then ϕ is bounded, i.e., there is some constant c_2 such that $|\phi(\rho)| \leq c_2$ for all $\rho \in [0,1]$. Moreover, if f is not the indicator of an $(n-k)$-subcube, it holds that (cf. Theorem 9.8)

$$\frac{\partial}{\partial \rho} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=1} > \frac{q}{2}\, a \,\log\left(\frac{1}{a}\right).$$

By again exploiting that fact that the number of Boolean functions is finite,

$$c_1 := \min_{f \in \mathcal{F}} \frac{\partial}{\partial \rho} \mathbf{S}_\rho^{(q)}[f] \Big|_{\rho=1} > \frac{q}{2}\, a \log\left(\frac{1}{a}\right),$$

where $\mathcal{F}$ denotes the set of Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that *cannot* be written as the indicator of an $(n-k)$-subcube. Therefore, for any $f \in \mathcal{F}$,

$$\mathbf{S}_\rho^{(q)}[f] \leq a + c_1(\rho - 1) + c_2(\rho - 1)^2. \tag{9.43}$$

By Taylor's theorem, one can lower bound the q-stability for the indicator of an $(n-k)$-subcube $\mathbb{C}_{n-k}$ as

$$\mathbf{S}_\rho^{(q)}[\mathbb{1}_{\mathbb{C}_{n-k}}] \geq a + (\rho - 1)\frac{q}{2}\, a \log\left(\frac{1}{a}\right) + c_3(\rho - 1)^2, \tag{9.44}$$

where c_3 is an absolute constant independent of ρ. Comparing (9.43) and (9.44), we observe that there exists a constant $\epsilon > 0$ such that the right-hand side of (9.44) is larger than (9.43) for all $\rho \in [1-\epsilon, 1]$, concluding the proof sketch of Theorem 9.9. □

Concerning the other extreme case, i.e., the limiting case as $\rho \downarrow 0$, following the proof ideas used in Theorems 9.8 and 9.9, one can also show the following result, which is due to Mossel and O'Donnell [122] and Li and Médard [110].

Theorem 9.10. Fix $n \geq 1$, $q > 1$, and a dyadic rational $a \in (0,1)$. There exists an $\epsilon \in (0,1)$ such that for all $\rho \in [0,\epsilon]$, $\Gamma_\rho^{(q)}(a)$ is attained by some Boolean function that maximizes the degree-1 Fourier weight $\mathbf{W}_1$. In particular, for $a = 1/2$, there exists an $\epsilon > 0$ such that for all $\rho \in [0,\epsilon]$, $\Gamma_\rho^{(q)}(1/2)$ is attained by dictator functions.

Theorems 9.8, 9.9, and 9.10 can be extended to their symmetric counterparts of the max q-stability. For $q = 1$, they can also be extended to the Φ-versions of the max q-stabilities (cf. Definition 9.6); see Courtade and Kumar [40], Ordentlich, Shayevitz, and Weinstein [134], and Yang and Wesel [187].

9.4 The Balanced Case

In this section, we consider the balanced case, i.e., $a = 1/2$, and discuss recent progress on Conjectures 9.1 and 9.2. We first focus on the case $q = 1$, i.e., the balanced version of the Courtade–Kumar conjecture, which can be stated as follows.

Conjecture 9.3. For any $n \in \mathbb{N}$ and $\rho \in (0, 1)$,

$$\max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=1/2} I(f(X^n); Y^n) \stackrel{?}{=} 1 - h\Big(\frac{1-\rho}{2}\Big). \tag{9.45}$$

In the original version of Courtade–Kumar conjecture, the Boolean function f is not required to satisfy $\mathbb{E}[f(X^n)] = 1/2$. It has been numerically verified to be true for all $n \le 7$ [40]. An old result by Witsenhausen and Wyner [176] (also see Erkip [53]) yields the following bound.

Proposition 9.4. It holds that

$$\max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=1/2} I(f(X^n); Y^n) \le \rho^2. \tag{9.46}$$

This proposition can be proved via the so-called Mrs. Gerber's lemma [184] or the hypercontractivity inequality in (8.56). Here, we provide a short justification based on the latter. By (8.56), we obtain that for $q > 1$ and any Boolean function f with mean a,

$$\mathbf{S}_\rho^{(q)}[f] \le a^{\frac{q}{1+(q-1)\rho^2}}.$$

In other words,

$$\Gamma_\rho^{(q)}(a) \le a^{\frac{q}{1+(q-1)\rho^2}} \quad \text{and} \quad \breve{\Gamma}_\rho^{(q)}(a) \le a^{\frac{q}{1+(q-1)\rho^2}} + \bar{a}^{\frac{q}{1+(q-1)\rho^2}}.$$

Substituting the latter into (9.25) and setting $a = 1/2$ yields

$$\breve{\Pi}_\rho^{(q)}(1/2) \le \frac{2^{\frac{(1-q)(1-\rho^2)}{1+(q-1)\rho^2}} - 1}{(q-1)\ln 2}.$$

Letting $q \downarrow 1$, we obtain $\breve{\Pi}_\rho^{(1)}(1/2) \le \rho^2 - 1$. Substituting this into (9.28) and noting that $h(1/2) = 1$ yields (9.46) as desired.

Considering small ρ, and using Fourier analysis and hypercontractivity, Ordentlich, Shayevitz, and Weinstein [134] improved the bound in (9.46) to the following.

Proposition 9.5. For $0 \le \rho \le 1/\sqrt{3}$,

$$\max_{\text{Boolean } f:\mathbb{E}[f(X^n)]=1/2} I(f(X^n);Y^n) \le \frac{\log e}{2}\rho^2 + 9\Big(1 - \frac{\log e}{2}\Big)\rho^4. \quad (9.47)$$

The bounds in (9.46) and (9.47) are illustrated in Fig. 9.6. The bound in (9.47) is better than (9.46) in the range $0 < \rho < 1/3$. Moreover, the bound in (9.47) is asymptotically tight as $\rho \downarrow 0$, i.e., the ratio of the bound in (9.47) and the right-hand side of (9.45) converges to 1 as $\rho \downarrow 0$. This point can be seen from the fact that by Taylor's theorem, as $\rho \downarrow 0$,

$$1 - h\Big(\frac{1-\rho}{2}\Big) = \frac{\log e}{2}\rho^2 + \frac{\log e}{12}\rho^4 + O(\rho^6).$$

In 2016, Samorodnitsky [149] made a significant breakthrough on the Courtade–Kumar conjecture. Specifically, he proved the existence of a *dimension-independent* interval for which Conjecture 9.3 holds for all ρ in the interval.

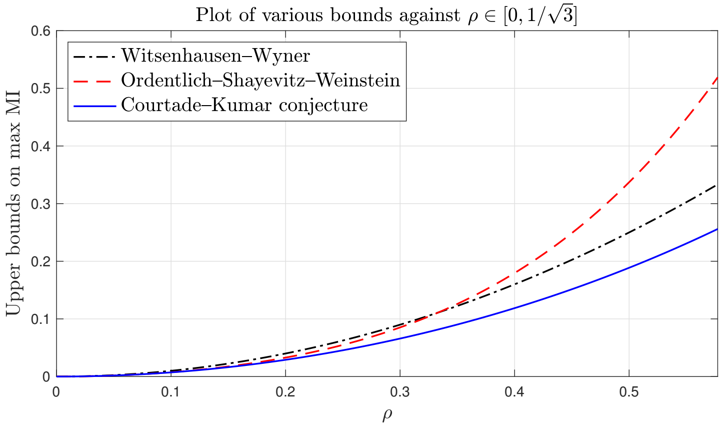

Figure 9.6: Illustration bounds on $\max I(f(X^n);Y^n)$ by Witsenhausen and Wyner [176] in (9.46), Ordentlich, Shayevitz, and Weinstein [134] in (9.47), and the Courtade–Kumar conjecture in (9.45)

Theorem 9.11. There exists a constant $0 < \rho_0 < 1$ (independent of n), such that (9.45) holds for any $n \in \mathbb{N}$ and any $\rho \in (0, \rho_0]$.

The proof by Samorodnitsky [149] is based on Fourier analysis, random restrictions, techniques in Ordentlich, Shayevitz, and Weinstein

[134], the Friedgut–Kalai–Naor (FKN) theorem [90], among others. Samorodnitsky's proof is highly technical, so we do not present it here. However, we should note that in the proof of Theorem 9.11, ρ_0, which is not explicitly provided, is assumed to be "sufficiently small". It is also worth noting that the conclusion that the value ρ_0 is independent of n (resulting in a *dimension-independent* interval $(0, \rho_0])$ is the crux of this theorem. Indeed, if we allow ρ_0 to vary with n, then the resulting theorem is merely an extension of Theorem 9.10 to the case $q = 1$, which can be proved by combining the bound in (9.47) by Ordentlich, Shayevitz, and Weinstein [134] and the discreteness of the space of Boolean functions; see [134, Corollary 1]. This fact can also be deduced using calculus [187].

Using Fourier analysis and optimization theory, the first author of this monograph [191] provided an explicit threshold for Theorem 9.11. Specifically, he showed that (9.45) holds for any n and any $\rho \in (0, \rho_1]$, where ρ_1 be the solution in $(0, 1)$ to the equation

$$(1 + \rho^2) \log\left(\frac{1+\rho}{2}\right) - (1-\rho)^2 \log\left(\frac{1-\rho}{2}\right) = 0.$$

The value of $\rho_1 \approx 0.461491$.

In the Courtade–Kumar conjecture, if the Boolean function is set to a dictator function $f(x^n) = x_1$ (say), then the objective function $I(f(X^n); Y^n) = I(X_1; Y_1)$. Motivated by this, in addition to the original Courtade–Kumar conjecture (in which f is an arbitrary Boolean function and not required to satisfy $\mathbb{E}[f(X^n)] = 1/2$), Courtade and Kumar also proposed a weaker version of this conjecture. They conjectured that for any $n \in \mathbb{N}$ and $\rho \in (0, 1)$,

$$\max_{\text{Boolean } f,g} I(f(X^n); g(Y^n)) = 1 - h\Big(\frac{1-\rho}{2}\Big). \tag{9.48}$$

This weaker version was proven by Pichler, Piantanida, and Matz [136] by using Fourier analysis and a novel partitioning technique.

Theorem 9.12. The equality in (9.48) holds for all $(n, \rho) \in \mathbb{N} \times (0, 1)$.

Since the Li–Médard conjecture was only recently posed (at the time of writing), there is less progress on it compared to the Courtade–Kumar conjecture. Hence, we do not elaborate on it apart from mentioning some partial progress by Yu [191] for a certain set of (q, ρ).

Finally, we summarize some recent progress on the Mossel–O'Donnell conjecture, which states that dictator functions are optimal in attaining both the asymmetric and symmetric max q-stabilities for $2 < q \leq 9$ (and for any $n \in \mathbb{N}$ and any $\rho \in (0, 1)$). As discussed in Theorems 9.9 and 9.10, the limiting cases as $\rho \downarrow 0$ and $\rho \uparrow 1$ (with fixed n) were resolved in [122] for the symmetric case and in [110], [122] for the asymmetric case. However, for other intermediate values of ρ, there has been fairly limited progress. For the symmetric case, the best known result is Mossel and O'Donnell's result in Theorem 9.2; this result resolved the eponymous conjecture for $q = 3$ and for any $\rho \in (0, 1)$. Combining this with the result of Barnes and Özgür [11] (in Lemma 9.3) yields the conclusion the Mossel–O'Donnell conjecture holds for all $2 < q \leq 3$. There is even less progress for the asymmetric case in which the best known result remains that of Witsenhausen's result in Corollary 9.1 for the case $q = 2$. Recently, in [191], the first author of this monograph made some progress on the Mossel–O'Donnell conjecture. He showed that the symmetric version of the Mossel–O'Donnell conjecture holds for $2 < q \leq 5$, and the asymmetric version holds for $2 < q \leq 3$. These imply that $3 \leq q_{\max} < 10$ and $5 \leq \breve{q}_{\max} < 10$. The proofs are based on Fourier analysis and optimization theory.

9.5 Moderate and Large Deviations Regimes

In this section, we consider the max q-stabilities in the MD and LD regimes. Recall the definition of the asymmetric max q-stability with $q \in [1, \infty)$ in (9.9). It can be rewritten as

$$\Gamma_\rho^{(q)}(a) = \left(\max_{\mathcal{A} \subset \{0,1\}^n : \pi_X^n(\mathcal{A}) \leq a} \|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q \right)^q, \tag{9.49}$$

where the maximization is over all subsets of $\{0, 1\}^n$. We now extend the asymmetric max q-stability to the case of $q \in (-\infty, 1)\backslash\{0\}$. For $q \in (-\infty, 1)\backslash\{0\}$, define

$$\Gamma_\rho^{(q)}(a) := \left(\min_{\mathcal{A} \subset \{0,1\}^n : \pi_X^n(\mathcal{A}) \geq a} \|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q \right)^q. \tag{9.50}$$

We note that even though a min is present in (9.50), we still term this quantity as the asymmetric *max* q-stability.

We are now interested in the MD and LD asymptotics of (9.49) and (9.50). Similarly to the 2-user NICD problem, in the LD regime, the parameter a is assumed to vanish exponentially fast as $n \to \infty$, i.e., $a = 2^{-n\alpha}$ for some fixed constant $\alpha \in (0, 1)$. In the MD regime, a is assumed to vanish subexponentially fast, i.e., $a = 2^{-\theta_n \alpha}$ for an MD sequence $\{\theta_n\}_{n\in\mathbb{N}}$.

Definition 9.11. We define the LD and MD exponents corresponding to the quantities in (9.49) and (9.50) as follows.

1. For $n \geq 1$, $\alpha \in [0, 1]$, and $q \geq 1$, define the *LD exponent* as

$$\Upsilon_{q,\mathrm{LD}}^{(n)}(\alpha) := -\frac{1}{n} \log \max_{\mathcal{A}:\pi_X^n(\mathcal{A}) \leq 2^{-n\alpha}} \|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q. \tag{9.51}$$

   For $q \in (-\infty, 1)\backslash\{0\}$, $\Upsilon_{q,\mathrm{LD}}^{(n)}(\alpha)$ is defined similarly but with the maximization in (9.51) replaced by a minimization, and the inequality reversed.

2. For $n \geq 1$, $\alpha \in [0, \infty)$, $q \geq 1$, and an MD sequence $\{\theta_n\}_{n\in\mathbb{N}}$, define the *MD exponent* as

$$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha) := -\frac{1}{\theta_n} \log \max_{\mathcal{A}:\pi_X^n(\mathcal{A}) \leq 2^{-\theta_n\alpha}} \|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q. \tag{9.52}$$

   For $q \in (-\infty, 1)\backslash\{0\}$, $\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha)$ is defined similarly but with the maximization in (9.52) replaced by a minimization, and the inequality reversed.

3. Define $\Upsilon_{q,\mathrm{MD}}^{(\infty)}$ and $\Upsilon_{q,\mathrm{LD}}^{(\infty)}$ as the pointwise limits of (9.51) and (9.52) as $n \to \infty$.

Note that in the definitions in (9.51)–(9.52), we remove the q^{th} power in (9.49)–(9.50). This slight modification will result in a multiplicative factor of q in the characterizations of these exponents. We deliberately choose such definitions since the bounds on the exponents in Definition 9.11 provided in the following two theorems will be consistent with the bounds for the 2-user NICD problem. We also remark that these quantities depend on ρ but these dependencies are suppressed to avoid notational clutter in what follows.

For $q \in (1-\rho^{-2}, \infty)\backslash\{0\}$ and $\alpha > 0$, let

$$\Upsilon_{q,\mathrm{MD}}(\alpha) := \frac{\alpha}{1+(q-1)\rho^2}. \tag{9.53}$$

By using the single-function versions of hypercontractivity inequalities (Theorem 8.5), we can obtain the following result.

Theorem 9.13 (q-stability)**.** Let $n \geq 1$ and $\alpha > 0$. For $q \geq 1$,

$$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha) \geq \Upsilon_{q,\mathrm{MD}}(\alpha), \tag{9.54}$$

and for $q \in (1-\rho^{-2}, 1)\backslash\{0\}$,

$$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha) \leq \Upsilon_{q,\mathrm{MD}}(\alpha). \tag{9.55}$$

Moreover, these two bounds are asymptotically tight, i.e., for $q \in (1-\rho^{-2}, \infty)\backslash\{0\}$,

$$\Upsilon_{q,\mathrm{MD}}^{(\infty)}(\alpha) = \Upsilon_{q,\mathrm{MD}}(\alpha). \tag{9.56}$$

Lastly, for $q \in (-\infty, 1-\rho^{-2}]$, $\Upsilon_{q,\mathrm{MD}}^{(\infty)}(\alpha) = \infty$. The equalities are achieved by sequences of Hamming balls or spherical shells.

The function $\Upsilon_{q,\mathrm{MD}}$, defined in (9.53), is plotted in Fig. 9.7.

Proof of Theorem 9.13. This theorem is a consequence of the classic hypercontractivity inequalities in (8.56) and (8.57). Substituting $f \leftarrow \mathbb{1}_{\mathcal{A}}$ into (8.56) and (8.57), we obtain for $q \geq 1$,

$$\|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q \leq \pi_X^n(\mathcal{A})^{\frac{1}{1+(q-1)\rho^2}},$$

and for $q \in (1-\rho^{-2}, 1)\backslash\{0\}$,

$$\|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q \geq \pi_X^n(\mathcal{A})^{\frac{1}{1+(q-1)\rho^2}}.$$

These inequalities immediate imply (9.54) and (9.55).

The asymptotic tightness of (9.54) and (9.55) can be verified by choosing the sets $\mathcal{A}$ in the definition of the MD exponent to be sequences of Hamming balls or spherical shells. The asymptotic tightness for $q \in (-\infty, 1-\rho^{-2}]$ follows by the monotonicity of the L^q-norm in q, and taking limits as $q \downarrow 1-\rho^{-2}$ in (9.56). We omit the details. $\square$

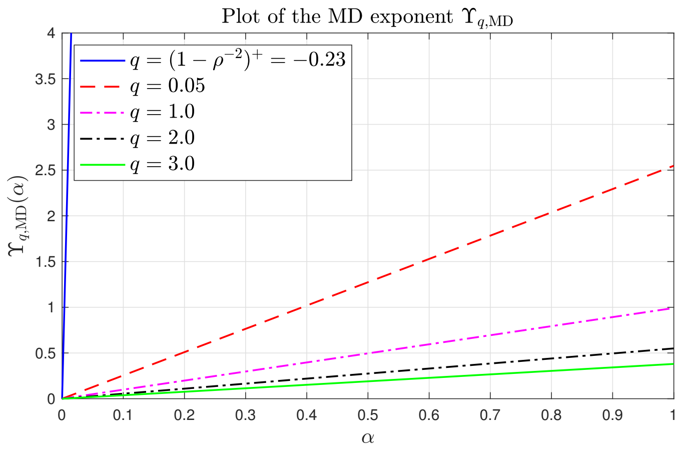

Figure 9.7: The MD exponent of the q-stability $\Upsilon_{q,\mathrm{MD}}$ for $\rho = 0.9$. Observe that $\Upsilon_{q,\mathrm{MD}}$ is linear given each $q \neq 0$ and diverges as $q \downarrow 1 - \rho^{-2} \approx -0.2346$.

We now turn our attention to the LD exponent. For $q \neq 0$, define

$$\theta_q(Q_X, Q_Y) := \mathsf{D}(Q_X, Q_Y \| \pi_{XY}) - \frac{D(Q_Y \| \pi_Y)}{q'}. \tag{9.57}$$

where q' is the Hölder conjugate of q Define

$$\Upsilon_{q,\mathrm{LD}}(\alpha) := \inf_{Q_X, Q_Y : D(Q_X \| \pi_X) \geq \alpha} \theta_q(Q_X, Q_Y) \tag{9.58}$$

for $q \geq 1$, and

$$\Upsilon_{q,\mathrm{LD}}(\alpha) := \begin{cases} \sup_{Q_X : D(Q_X \| \pi_X) \leq \alpha} \inf_{Q_Y} \theta_q(Q_X, Q_Y) & 0 < q < 1 \\ \sup_{Q_X : D(Q_X \| \pi_X) \leq \alpha} \sup_{Q_Y} \theta_q(Q_X, Q_Y) & q < 0 \end{cases} \tag{9.59}$$

for $q \in (-\infty, 1) \backslash \{0\}$. It can be verified that $\Upsilon_{q,\mathrm{LD}}(\alpha) \geq 0$ for all $q \neq 0$. Asymptotically tight bounds are provided in the following theorem, which is known as the *strong q-stability theorem* and was proved by the first author of this monograph [192].

Theorem 9.14 (Strong q-stability)**.** For any $n \geq 1$ and $\alpha \in (0, 1)$, it holds that for $q \geq 1$,

$$\Upsilon^{(n)}_{q,\mathrm{LD}}(\alpha) \geq \mathbb{L}[\Upsilon_{q,\mathrm{LD}}](\alpha), \tag{9.60}$$

and for $q \in (-\infty, 1)\backslash\{0\}$,

$$\Upsilon^{(n)}_{q,\mathrm{LD}}(\alpha) \le \mathbb{U}[\Upsilon_{q,\mathrm{LD}}](\alpha). \tag{9.61}$$

Moreover, these two bounds are asymptotically tight, i.e.,

$$\Upsilon^{(\infty)}_{q,\mathrm{LD}}(\alpha) = \mathbb{L}[\Upsilon_{q,\mathrm{LD}}](\alpha) \quad \text{and} \quad \Upsilon^{(\infty)}_{q,\mathrm{LD}}(\alpha) = \mathbb{U}[\Upsilon_{q,\mathrm{LD}}](\alpha), \tag{9.62}$$

and these equalities are achieved by sequences of Hamming balls or spheres.

It has been shown in [193] that for $q \ge 1$, $\Upsilon_{q,\mathrm{LD}}$ is convex, and for $q \in (-\infty, 1)\backslash\{0\}$, $\Upsilon_{q,\mathrm{LD}}$ is concave. Combining this result with the strong q-stability theorem (Theorem 9.14) tells us that the lower convex envelope and upper concave envelope operations in (9.62) can be removed and Hamming balls or spheres are optimal in the LD regime. That is, for the DSBS and $\alpha \in (0, 1)$,

$$\Upsilon^{(\infty)}_{q,\mathrm{LD}}(\alpha) = \Upsilon_{q,\mathrm{LD}}(\alpha) \quad \text{and} \quad \Upsilon^{(\infty)}_{q,\mathrm{LD}}(\alpha) = \Upsilon_{q,\mathrm{LD}}(\alpha).$$

This is parallel to the discussion of the resolution of the OPS conjecture in Section 8.5; also see (8.67). The asymptotically tight bound $\Upsilon_{q,\mathrm{LD}}$, defined in (9.59), is plotted in Fig. 9.8 for various q's.

9.6 Extensions to Sources Beyond the DSBS

Similarly to the discussion in Section 8.6, the q-stability and strong q-stability theorems can be extended to sources defined on arbitrary finite alphabets as well as jointly Gaussian sources. We discuss these extensions here.

9.6.1 Finite Alphabets

Let π_{XY} be a joint distribution defined on a finite alphabet. We now consider its q-stability. For $\alpha \in [0, \alpha_{\max}(\pi_X)]$ (defined in (8.68)), we reuse the definitions in (9.51)–(9.52) for $\Upsilon^{(n)}_{q,\mathrm{MD}}$ and $\Upsilon^{(n)}_{q,\mathrm{LD}}$, but with the underlying distribution set to be π_{XY}. The strong q-stability theorem (Theorem 9.13) can be extended to the following general version, which was first shown in [192], as a consequence of the strong version of hypercontractivity inequalities derived in [192]. We provide a simple proof of Theorem 9.15 in Section 10.3.

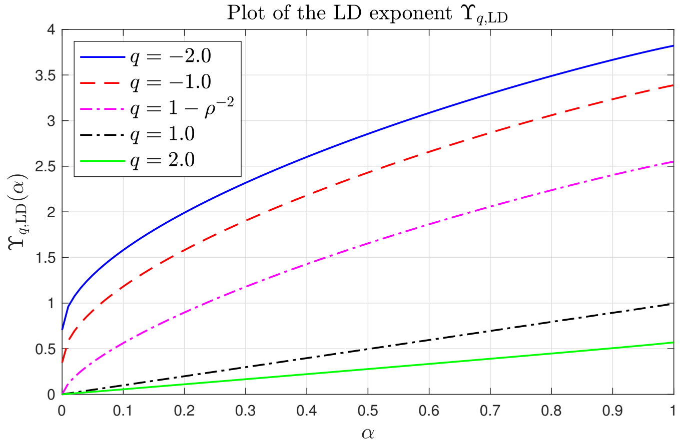

Figure 9.8: The LD exponent of the q-stability $\Upsilon_{q,\mathrm{LD}}$ for $\rho = 0.9$. Observe that $\Upsilon_{q,\mathrm{LD}}$ appears to be ("only slightly") convex for $q > 1$, concave for $q \in (-\infty, 1) \setminus \{0\}$, and linear for $q = 1$. Also $\Upsilon_{q,\mathrm{LD}}(0)$ vanishes when $q \leq 1 - \rho^{-2} \approx -0.2346$.

Theorem 9.15 (Strong q-stability: General version)**.** For any $n \geq 1$ and $\alpha \in (0, \alpha_{\max}(\pi_X)]$, (9.60) holds for $q \geq 1$, and (9.61) holds for $q \in (-\infty, 1) \setminus \{0\}$. Moreover, (9.60) and (9.61) are asymptotically tight, i.e., (9.62) holds.

The q-stability theorem (Theorem 9.14) can be also generalized to the finite alphabet case, but for general sources on finite alphabets, a limiting operation is needed.

Theorem 9.16 (q-Stability: General version)**.** For any $n \geq 1$, $\alpha > 0$, and $q \geq 1$,

$$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha) \geq \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbb{L}[\Upsilon_{q,\mathrm{LD}}](\epsilon\alpha). \tag{9.63}$$

If instead $q \in (-\infty, 1) \setminus \{0\}$ and $\Upsilon_{q,\mathrm{LD}}(0) = 0$, then

$$\Upsilon_{q,\mathrm{MD}}^{(n)}(\alpha) \leq \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \mathbb{U}[\Upsilon_{q,\mathrm{LD}}](\epsilon\alpha). \tag{9.64}$$

Moreover, the inequalities in (9.63)–(9.64) are asymptotically tight.

9.6.2 Gaussian Sources

Finally, we turn our attention to memoryless bivariate Gaussian sources with correlation coefficient $\rho \in (0, 1)$. For this class of sources, the max q-stability problem was completely solved by Borell [28] for all $a \in [0,1]$. Let π_{XY} be a bivariate Gaussian distribution with zero mean and covariance matrix $\mathbf{K}$ given in (8.17). Let $(X^n, Y^n) \sim \pi_{XY}^n$. For this distribution, a real number $q > 1$, and $a \in [0,1]$, we define

$$\Gamma_\rho^{(q)}(a) := \sup \|\pi_{X|Y}^n(\mathcal{A}|Y^n)\|_q^q,$$

where the supremum runs over all measurable sets $\mathcal{A} \subset \mathbb{R}^n$ such that $\pi_X^n(\mathcal{A}) = a$. The following theorem is due to Borell [28].

Theorem 9.17 (Borell's q-stability theorem)**.** For any $n \geq 1$, $q > 1$, $0 \leq \rho < 1$, and $a \in [0,1]$, one has

$$\Gamma_\rho^{(q)}(a) = \Lambda_\rho^{(q)}(a), \tag{9.65}$$

where $\Lambda_\rho^{(q)}$, the Gaussian q-stability function, is defined in (9.6). Moreover, optimal subsets $\mathcal{A}$ (i.e., those attaining $\Gamma_\rho^{(q)}$) are parallel halfspaces.

The proof of this theorem can be found in Borell [28] and Eldan [52]. Moreover, the proof of this theorem with q being an integer can also be found in Isaksson and Mossel [87] and Neeman [129]. We remark that Neeman's proof in [129] is an extension of the one for Borell's isoperimetric theorem given in Section 8.6.2 to the multi-user case.

We now consider the Gaussian version of the Courtade–Kumar conjecture. Substituting (9.65) into the Φ-symmetric max q-stability in (9.24) and taking limits as $q \downarrow 1$, one can deduce that $\breve{\Pi}_\rho^{(1)}(a)$ is attained by halfspaces with π_X^n-probability a. This implies that

$$\begin{aligned}&\max_{\substack{f:\mathbb{R}^n\to\{0,1\}\text{ measurable:}\\ \mathbb{E}[f(X^n)]=a}} -H(f(X^n)|Y^n)\\ &= -H(\mathbb{1}\{X_1 \leq \Phi^{-1}(a)\}|Y_1) = -\mathbb{E}_{Y_1}\left[h\left(\Phi\left(\frac{\Phi^{-1}(a)-\rho Y_1}{\sqrt{1-\rho^2}}\right)\right)\right].\end{aligned} \tag{9.66}$$

That is, given $a \in [0,1]$, the indicator of any half-space with π_X^n-probability a (e.g., $(-\infty, \Phi^{-1}(a)] \times \mathbb{R}^{n-1}$) maximizes the mutual information between $f(X^n)$ and Y^n over all $\{0,1\}$-valued measurable functions f. This statement was also proved by Kindler, O'Donnell, and Witmer [96] using a different method. If we do not fix a, then similarly to the original Courtade–Kumar conjecture for the DSBS, it is natural to conjecture that for this Gaussian version of Courtade–Kumar conjecture, the mutual information is also maximized at $a = 1/2$ for every $\rho \in (0, 1)$. This point can be confirmed numerically as shown in Fig. 9.9 in which we plot the right-hand side of (9.66) plus $h(a)$ as a function of $a \in [0, 1/2]$ for different ρ's. Note that we only focus on the case $a \in [0, 1/2]$ in Fig. 9.9, since the function considered is symmetric with respect to $a = 1/2$. It is easily seen that the maxima of these curves occur at $a = 1/2$.

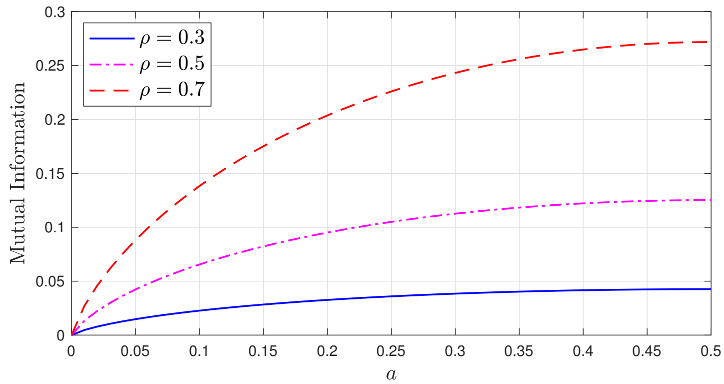

Figure 9.9: The mutual information, i.e., the right-hand side of (9.66) plus $h(a)$.

10

Functional Inequalities

In this section, we consider functional extensions of the NICD and the max q-stability problems as described in Sections 8 and 9 respectively. Recall that in the 2-user NICD problem, we optimize the probability of agreement between two random bits that are generated in a distributed manner via *Boolean functions* from a joint source (X^n, Y^n). In this section, we replace the Boolean functions $f, g \in \{0,1\}^n \to \{0,1\}$ with *arbitrary nonnegative functions*, and obtain corresponding functional inequalities. Specifically, we will introduce the Brascamp–Lieb inequalities, the hypercontractivity inequalities, and the log-Sobolev inequalities, as well as their strengthened counterparts. We provide information-theoretic characterizations of these inequalities, and also use them to prove the strong SSE theorem and the strong q-stability theorem stated respectively in Sections 8.6 and 9.6. Analogously to the forward and reverse joint probabilities in the NICD problem (Definition 8.1), the optimal constants or exponents in these inequalities can be also regarded as refinements of GKW's common information when the latter is equal to zero, but with the "information" measured by the *entropy* of a nonnegative function, rather than the Shannon entropy.

This section concerning *functional inequalities* (or inequalities involving functionals) starts by formally defining some convenient quantities, such as the minimum relative entropy region, in Section 10.1. Using these new definitions, we provide alternative representations of the forward and reverse large deviations exponents in the NICD and q-stability problems. These quantities are then used in Section 10.2 to express the hypercontractivity regions (which generalize and strengthen the classic Hölder inequalities) and Brascamp–Lieb exponents in terms of single-letter, information-theoretic quantities. We then connect these exponents to the NICD and q-stability problems in Section 10.3, leading to a short proof of the strong SSE theorem (Theorem 8.11). In Section 10.4, we discuss the log-Sobolev inequalities, provide single-letter expressions for their optimal constants, and use the results as a bridge to connect the hypercontractivity inequalities to their strengthened counterparts, which are presented in Section 10.5. In Section 10.5, our discussion culminates with expressions for the strong log-Sobolev constant and a strengthened hypercontractivity inequality for the DSBS. Throughout this section, we focus on *information-theoretic* characterizations of optimal constants and exponents in various functional inequalities.

As there are several interconnected results in this section and Sections 8 and 9, we illustrate their relationships by means of a graph in Fig. 10.1.

10.1 Preliminary Definitions

Throughout this section, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite sets and π_{XY} is a joint distribution on $\mathcal{X} \times \mathcal{Y}$.

Assumption 10.1 (Full support of marginals)**.** The supports of π_X and π_Y are $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Definition 10.1. Define the *minimum relative entropy region* with respect to a joint distribution $\pi_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ as

$$\mathcal{D}(\pi_{XY}) := \bigcup_{Q_X, Q_Y} \Big\{ (D(Q_X \| \pi_X), D(Q_Y \| \pi_Y), \mathsf{D}(Q_X, Q_Y \| \pi_{XY})) \Big\},$$

where $\mathsf{D}(Q_X, Q_Y \| \pi_{XY})$ is the minimal relative entropy with respect to π_{XY} over all couplings of Q_X and Q_Y, defined in (8.23). Due to

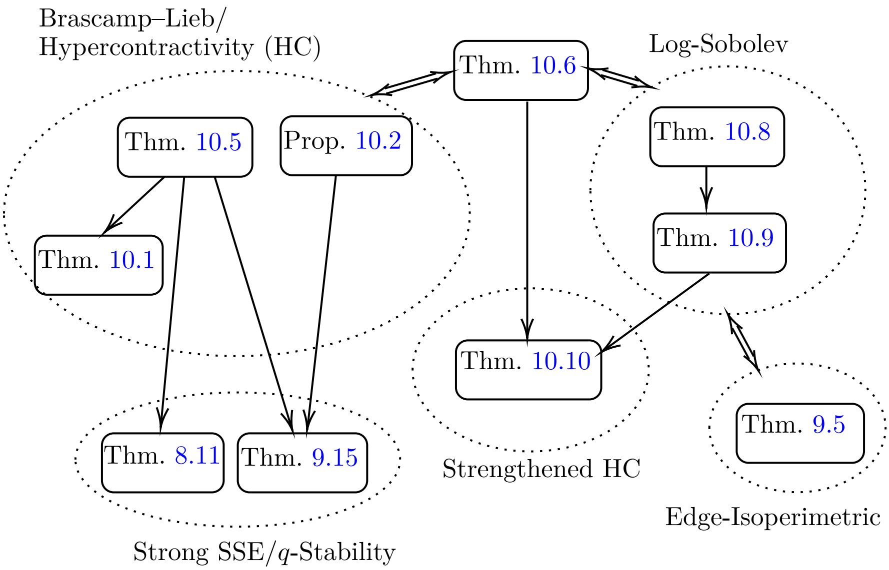

Figure 10.1: A graph of the main results in this and Sections 8 and 9, where $\longrightarrow$ denotes an implication and $\Longleftrightarrow$ denotes a close relationship.

Assumption 10.1, any Q_X and Q_Y defined on $\mathcal{X}$ and $\mathcal{Y}$ respectively are absolutely continuous with respect to π_X and π_Y respectively.

The minimum relative entropy region is the subset of $\mathbb{R}^3$ that is formed by the pair of relative entropies $(D(Q_X\|\pi_X), D(Q_Y\|\pi_Y))$ and the minimal relative entropy $\mathsf{D}(Q_X, Q_Y\|\pi_{XY})$ as Q_X and Q_Y run over all distributions that are absolutely continuous with respect to π_X and π_Y respectively.

Definition 10.2. For $(s,t) \in [0, \alpha_{\max}(\pi_X)] \times [0, \beta_{\max}(\pi_Y)]$ (refer to (8.68) for definitions), define the *upper* and *lower envelopes* of the minimal relative entropy region $\mathcal{D}(\pi_{XY})$ respectively as

$$\underline{\varphi}(s,t) := \min_{Q_X, Q_Y : D(Q_X\|\pi_X)=s, D(Q_Y\|\pi_Y)=t} \mathsf{D}(Q_X, Q_Y\|\pi_{XY}), \quad (10.1)$$

and

$$\overline{\varphi}(s,t) := \max_{Q_X, Q_Y : D(Q_X\|\pi_X)=s, D(Q_Y\|\pi_Y)=t} \mathsf{D}(Q_X, Q_Y\|\pi_{XY}). \quad (10.2)$$

Fix $(\alpha, \beta) \in [0, \alpha_{\max}(\pi_X)] \times [0, \beta_{\max}(\pi_Y)]$. Recall that the upper bound on the forward LD exponent, previously defined in (8.24), is

$$\underline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) = \min_{Q_X, Q_Y : D(Q_X \| \pi_X) \geq \alpha, D(Q_Y \| \pi_Y) \geq \beta} \mathsf{D}(Q_X, Q_Y \| \pi_{XY}), \quad (10.3)$$

and the lower bound on the reverse LD exponent, is

$$\overline{\Upsilon}_{\mathrm{LD}}(\alpha, \beta) = \max_{Q_X, Q_Y : D(Q_X \| \pi_X) \leq \alpha, D(Q_Y \| \pi_Y) \leq \beta} \mathsf{D}(Q_X, Q_Y \| \pi_{XY}). \quad (10.4)$$

Based on the functions presented in Definition 10.2, we may modify the definitions of $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ for the DSBS to a source $(X, Y) \sim \pi_{XY}$ defined on a finite alphabet as follows

$$\underline{\Upsilon}(\alpha, \beta) := \min_{s \geq \alpha, t \geq \beta} \mathbb{L}[\underline{\varphi}](s, t) \quad \text{and} \quad (10.5)$$

$$\overline{\Upsilon}(\alpha, \beta) := \max_{s \leq \alpha, t \leq \beta} \mathbb{U}[\overline{\varphi}](s, t). \quad (10.6)$$

Note that $\underline{\Upsilon}_{\mathrm{LD}}$ and $\overline{\Upsilon}_{\mathrm{LD}}$ in (10.3) and (10.4) may not be convex and concave respectively for arbitrary π_{XY} (see the discussion following Theorem 8.11). Hence, in the modified definitions in (10.5) and (10.6), we take the lower convex envelope for $\underline{\varphi}$ and the upper concave envelope for $\overline{\varphi}$. With these operations, $\underline{\Upsilon}(\alpha, \beta)$ is convex and nondecreasing in (α, β), and $\overline{\Upsilon}(\alpha, \beta)$ is concave and nondecreasing in (α, β).[1] Henceforth, we omit the subscript LD in (10.5) and (10.6).

Before presenting the next definition, we recall the definition of θ_q in (9.57) as

$$\theta_q(Q_X, Q_Y) := \mathsf{D}(Q_X, Q_Y \| \pi_{XY}) - \frac{D(Q_Y \| \pi_Y)}{q'},$$

but now, instead of being a DSBS, π_{XY} is an arbitrary distribution defined on the finite set $\mathcal{X} \times \mathcal{Y}$.

Definition 10.3. For $q \geq 1$ and for $s \in [0, \alpha_{\max}(\pi_X)]$, define

$$\varphi_q(s) := \min_{Q_X, Q_Y : D(Q_X \| \pi_X) = s} \theta_q(Q_X, Q_Y),$$

[1]We say a function of two variables is *nondecreasing* if it is nondecreasing in one argument when the other is fixed.

and for $q \in (-\infty, 1)\backslash\{0\}$, define

$$\varphi_q(s) := \begin{cases} \max_{Q_X : D(Q_X \| \pi_X) = s} \min_{Q_Y} \theta_q(Q_X, Q_Y) & 0 < q < 1 \\ \max_{Q_X : D(Q_X \| \pi_X) = s} \max_{Q_Y} \theta_q(Q_X, Q_Y) & q < 0 \end{cases}.$$

We denote Υ_q as the lower convex envelope of φ_q for $q \geq 1$ and the upper concave envelope of φ_q for $q \in (-\infty, 1)\backslash\{0\}$. Specifically, for $\alpha \in [0, \alpha_{\max}(\pi_X)]$,

$$\Upsilon_q(\alpha) := \begin{cases} \min_{s \geq \alpha} \mathbb{L}[\varphi_q](s) & q \geq 1 \\ \max_{s \leq \alpha} \mathbb{U}[\varphi_q](s) & q \in (-\infty, 1)\backslash\{0\} \end{cases}.$$

Observe that Υ_q is an alternative representation of $\Upsilon_{q,\mathrm{LD}}$ defined in (9.58) and (9.59). By definition, $\Upsilon_q(\alpha)$ is convex and nondecreasing in α for each $q \geq 1$, and concave and nondecreasing in α for each $q \in (-\infty, 1)\backslash\{0\}$.

To avoid having to deal with the undefined arithmetic operation $\infty - \infty$, we adopt the following convention.

Convention 10.1. When we write an optimization problem with distributions as decision variables, we implicitly require that the distributions satisfy the condition that all the integrals and relative entropies (appearing in the constraints and the objective function) to be *finite*. Otherwise, the value of the optimization problem is set to $+\infty$ if it is an infimization, and $-\infty$ if it is a supremization.

To keep notation uncluttered, we also adopt the following convention.

Convention 10.2. When we write an optimization over functions f and g, we implicitly require these functions to be *nonnegative*.

10.2 Classic Hypercontractivity and Brascamp–Lieb Inequalities

In this section, we introduce a class of functional inequalities, known as *Brascamp–Lieb (BL) inequalities*. We also review the well-known Hölder and hypercontractivity inequalities which are special cases of the BL inequalities. We introduced the hypercontractivity inequalities in the context of of the DSBS in Section 8.3.1. In contrast, here we study these inequalities for *arbitrary* sources defined on finite alphabets.

10.2.1 Hölder and Hypercontractivity Inequalities

We review the well-known *forward* and *reverse Hölder inequalities* here. Given a joint distribution π_{XY} and an extended real number $p \in \mathbb{R} \cup \{\pm\infty\}$, for any pair of nonnegative functions (f, g), the forward and reverse Hölder inequalities are respectively

$$\langle f, g\rangle \leq \|f\|_p \|g\|_q \quad \text{if} \quad p \geq 1 \quad \text{and} \tag{10.7}$$

$$\langle f, g\rangle \geq \|f\|_p \|g\|_q \quad \text{if} \quad p \leq 1, \tag{10.8}$$

where q is the Hölder conjugate of p. Since the (pseudo) L^q-norms $\|\cdot\|_q$ are nondecreasing in $q \in \mathbb{R} \cup \{\pm\infty\}$, the scalar q in (10.7) can be replaced by any $q \geq p'$. Similarly, q in (10.8) can be replaced by any $q \leq p'$.

If $X = Y$ (i.e., $P_{Y|X}(\cdot|x)$ places all its mass at x for every $x \in \mathcal{X}$), then the forward and reverse Hölder inequalities are sharp in the following sense. If $p > 1$, then (10.7) becomes an equality if and only if $|f|^p$ and $g^{p'}$ are *linearly dependent*, i.e., there exist real numbers $a, b \geq 0$, not both zero, such that $a|f|^p = b|g|^{p'}$ holds (π_X-almost everywhere). If $p < 1$, $\langle f, g\rangle < \infty$ and $\|g\|_{p'} > 0$, then (10.8) is an equality if and only if the equality $|f|^p = a|g|^{p'}$ holds (π_X-almost everywhere) for some $a \geq 0$. Moreover, for the case $X = Y$, the parameters (p, q) in (10.7) and (10.8) cannot be improved in the sense that given $p \geq 1$, for any $q < p'$, there exists a pair of (f, g) that violates (10.7); similarly, given $p \leq 1$, for any $q > p'$, there exists a pair of (f, g) that violates (10.8).

However, the Hölder inequalities are not sharp in general when $X \neq Y$ (which is the case of interest to us). If $X \neq Y$, then the parameters (p, q) in the Hölder inequalities can be "improved". Specifically, given a joint distribution π_{XY} and $p \geq 1$, we are interested in how *small* $q \in \mathbb{R} \cup \{\pm\infty\}$ can be such that for *any* nonnegative functions $f : \mathcal{X} \to [0, \infty)$ and $g : \mathcal{Y} \to [0, \infty)$, it holds that

$$\langle f, g\rangle \leq \|f\|_p \|g\|_q. \tag{10.9}$$

By the forward Hölder inequality, the infimum of all such q's is at most p', the Hölder conjugate of p. Similarly, given $p \leq 1$, we are interested in how *large* $q \in \mathbb{R} \cup \{\pm\infty\}$ can be such that for any nonnegative functions f and g, it holds that

$$\langle f, g \rangle \geq \|f\|_p \|g\|_q. \tag{10.10}$$

For this case, the supremum of all such q's is at least p'. Inequalities (10.9) and (10.10) for the case $X \neq Y$ are respectively termed the *forward* and *reverse hypercontractivity inequalities*, since the forward and reverse Hölder inequalities in (10.7) and (10.8) respectively are regarded as the (usual) *contractivity* inequalities, and inequalities (10.9) and (10.10) with improved (p, q) are *strengthenings* of the forward and reverse Hölder inequalities.

Inequalities (10.9) and (10.10) motivate the following definitions.

Definition 10.4. The *forward* and *reverse hypercontractivity regions* [15], [112] are respectively defined as

$$\mathcal{R}_{\mathrm{FH}}(\pi_{XY}) := \{(p, q) \in [1, \infty)^2 : \langle f, g \rangle \leq \|f\|_p \|g\|_q,\ \forall\, f, g \geq 0\}$$

and

$$\mathcal{R}_{\mathrm{RH}}(\pi_{XY}) := \{(p, q) \in (-\infty, 1]^2 : \langle f, g \rangle \geq \|f\|_p \|g\|_q,\ \forall\, f, g \geq 0\}.$$

By definition, these two regions correspond to the sets of parameters (p, q) for which the forward or reverse hypercontractivity inequalities in (10.9) and (10.10) hold. We remark that the notion of *hypercontractivity ribbons* was introduced in Anantharam *et al.* [5, Eqn. (6.117)] and Kamath [92], prior to the hypercontractivity regions being introduced in Beigi and Gohari [15] and Liu [112]. The hypercontractivity ribbons correspond to the hypercontractivity regions apart from the exclusions of the Hölder regions $\{(p, q) \in [1, \infty)^2 : q \geq p'\}$ and $\{(p, q) \in (-\infty, 1]^2 : q \leq p'\}$, and that the Hölder conjugate of q is taken.

We can write $\mathcal{R}_{\mathrm{RH}}(\pi_{XY})$ as the disjoint union of four sets

$$\mathcal{R}_{\mathrm{RH}}^{++}(\pi_{XY}) := (0, 1]^2 \cap \mathcal{R}_{\mathrm{RH}}(\pi_{XY}), \tag{10.11}$$
$$\mathcal{R}_{\mathrm{RH}}^{+-}(\pi_{XY}) := ((0, 1] \times (-\infty, 0)) \cap \mathcal{R}_{\mathrm{RH}}(\pi_{XY}), \tag{10.12}$$
$$\mathcal{R}_{\mathrm{RH}}^{-+}(\pi_{XY}) := ((-\infty, 0) \times (0, 1]) \cap \mathcal{R}_{\mathrm{RH}}(\pi_{XY}), \quad \text{and} \tag{10.13}$$
$$\mathcal{R}_{\mathrm{RH}}^{--}(\pi_{XY}) := (-\infty, 0]^2.$$

The forward hypercontractivity region and the first three subregions of the reverse hypercontractivity region in (10.11), (10.12), and (10.13) admit the following information-theoretic characterizations; see [1], [16], [33], [92], [112], [192].

Theorem 10.1 (Information-theoretic characterizations of hypercontractivity regions). The forward hypercontractivity region $\mathcal{R}_{\mathrm{FH}}(\pi_{XY})$ can be expressed in terms of the minimal relative entropy as the set of $(p, q) \in [1, \infty)^2$ such that

$$\mathsf{D}(Q_X, Q_Y \| \pi_{XY}) \geq \frac{1}{p} D(Q_X \| \pi_X) + \frac{1}{q} D(Q_Y \| \pi_Y).$$

In addition, $\mathcal{R}_{\mathrm{RH}}^{++}(\pi_{XY})$ is the set of all $(p, q) \in (0, 1]^2$ such that

$$\mathsf{D}(Q_X, Q_Y \| \pi_{XY}) \leq \frac{1}{p} D(Q_X \| \pi_X) + \frac{1}{q} D(Q_Y \| \pi_Y).$$

Finally, $\mathcal{R}_{\mathrm{RH}}^{+-}(\pi_{XY})$ is the set of all $(p, q) \in (0, 1] \times (-\infty, 0)$ such that

$$\min_{Q_Y} \left\{ \mathsf{D}(Q_X, Q_Y \| \pi_{XY}) - \frac{1}{q} D(Q_Y \| \pi_Y) \right\} \leq \frac{1}{p} (Q_X \| \pi_X). \tag{10.14}$$

By symmetry, $\mathcal{R}_{\mathrm{RH}}^{-+}(\pi_{XY})$ can be characterized in an analogous manner to $\mathcal{R}_{\mathrm{RH}}^{+-}(\pi_{XY})$ in (10.14). The proof of Theorem 10.1 is provided in Section 10.2.2, since it is a special case of the information-theoretic characterizations of the BL inequalities, which we present therein. Theorem 10.1 can be specialized to Theorem 8.4 (the two function version of the hypercontractivity inequalities for the DSBS); see [127], [128].

Hypercontractivity inequalities were investigated in [1], [25]–[27], [69], [95], [125], [152] among others. Information-theoretic characterizations of the hypercontractivity (and BL) inequalities can be traced back to the seminal work of Ahlswede and Gács [1] in which, instead of the hypercontractivity regions, the hypercontractivity *constants* (which are quantities induced by the hypercontractivity regions) were characterized in terms of relative entropies. The information-theoretic characterization of the forward hypercontractivity region is implied by the information-theoretic characterization of the forward BL inequalities on Euclidean spaces in Carlen and Cordero-Erausquin [33]; this was independently discovered later by Nair [126] in the case of finite alphabets.

An information-theoretic characterization of $\mathcal{R}_{\mathrm{RH}}^{++}(\pi_{XY})$ for finite alphabets was provided by Kamath [92]. Subsequently, an information-theoretic characterization of the entire reverse hypercontractivity region for finite alphabets was shown by Beigi and Nair [16]. Extensions of these characterizations to Polish spaces were studied by Liu [112] using a minimax theorem known as the *Fenchel–Rockafellar duality*.

As a consequence of Definitions 10.2, 10.3, and Theorem 10.1, the regions $\mathcal{R}_{\mathrm{FH}}(\pi_{XY})$, $\mathcal{R}_{\mathrm{RH}}^{++}(\pi_{XY})$, and $\mathcal{R}_{\mathrm{RH}}^{+-}(\pi_{XY})$ also admit the following equivalent characterizations:

$$\begin{aligned}\mathcal{R}_{\mathrm{FH}}(\pi_{XY}) &= \Big\{(p,q) \in [1,\infty)^2 : \underline{\varphi}(\alpha,\beta) \geq \frac{\alpha}{p} + \frac{\beta}{q},\, \forall\, \alpha,\beta \geq 0\Big\} \\ &= \Big\{(p,q) \in [1,\infty)^2 : \underline{\Upsilon}(\alpha,\beta) \geq \frac{\alpha}{p} + \frac{\beta}{q},\, \forall\, \alpha,\beta \geq 0\Big\}, \\ \mathcal{R}_{\mathrm{RH}}^{++}(\pi_{XY}) &= \Big\{(p,q) \in (0,1]^2 : \overline{\varphi}(\alpha,\beta) \leq \frac{\alpha}{p} + \frac{\beta}{q},\, \forall\, \alpha,\beta \geq 0\Big\} \\ &= \Big\{(p,q) \in (0,1]^2 : \overline{\Upsilon}(\alpha,\beta) \leq \frac{\alpha}{p} + \frac{\beta}{q},\, \forall\, \alpha,\beta \geq 0\Big\},\end{aligned}$$

and

$$\begin{aligned}\mathcal{R}_{\mathrm{RH}}^{+-}(\pi_{XY}) &= \Big\{(p,q) \in (0,1] \times (-\infty,0) : \varphi_{q'}(\alpha) \leq \frac{\alpha}{p},\, \forall\, \alpha \geq 0\Big\} \\ &= \Big\{(p,q) \in (0,1] \times (-\infty,0) : \Upsilon_{q'}(\alpha) \leq \frac{\alpha}{p},\, \forall\, \alpha \geq 0\Big\},\end{aligned}$$

where q' is the Hölder conjugate of q.

10.2.2 Brascamp–Lieb Inequalities

The Brascamp–Lieb (BL) inequalities constitute a class of inequalities that generalizes the families of Hölder and hypercontractivity inequalities. The *forward* and *reverse BL inequalities* are defined as follows. Given a distribution π_{XY} and $p,q \in \mathbb{R}$, for any pair of nonnegative functions $f : \mathcal{X} \to [0,\infty)$ and $g : \mathcal{Y} \to [0,\infty)$,

$$\langle f,g\rangle \leq \overline{C}\,\|f\|_p\|g\|_q \quad \text{and} \tag{10.15}$$

$$\langle f,g\rangle \geq \underline{C}\,\|f\|_p\|g\|_q, \tag{10.16}$$

where $\overline{C} = \overline{C}_{p,q}$ and $\underline{C} = \underline{C}_{p,q}$ depend only on p and q given the distribution π_{XY}. The hypercontractivity inequalities in (10.9) and (10.10) correspond to the BL inequalities with $\overline{C} = 1$ in (10.15) and $\underline{C} = 1$ in (10.16) respectively.

The forward version of the BL inequalities in (10.15) was originally studied in the 1970s by Brascamp and Lieb [29], who were motivated by problems in particle physics. The reverse version in (10.16) was initially studied by Barthe [12]. In fact, the inequalities in (10.15) and (10.16) are special cases of the original forward and reverse BL inequalities. We only discuss these special cases.

Definition 10.5. The (optimal) *forward* and *reverse BL constants* are respectively defined as

$$\overline{C}^*_{p,q}(X;Y) := \sup_{f,g:\|f\|_p\|g\|_q>0} \frac{\langle f,g\rangle}{\|f\|_p\|g\|_q} \quad \text{and}$$
$$\underline{C}^*_{p,q}(X;Y) := \inf_{f,g:\|f\|_p\|g\|_q>0} \frac{\langle f,g\rangle}{\|f\|_p\|g\|_q}.$$

Additionally, define the *forward* and *reverse BL exponents* respectively as

$$\underline{\Lambda}_{p,q}(X;Y) := -\log \overline{C}^*_{p,q}(X;Y) \quad \text{and}$$
$$\overline{\Lambda}_{p,q}(X;Y) := -\log \underline{C}^*_{p,q}(X;Y).$$

It is well-known that the forward and reverse BL exponents possess the important tensorization and the data processing properties.

Lemma 10.2 (Tensorization)**.** Let $(X^n,Y^n) = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ be a collection of pairs of random variables that are mutually independent. Then

$$\underline{\Lambda}_{p,q}(X^n;Y^n) = \sum_{i=1}^n \underline{\Lambda}_{p,q}(X_i;Y_i) \quad \text{and} \tag{10.17}$$
$$\overline{\Lambda}_{p,q}(X^n;Y^n) = \sum_{i=1}^n \overline{\Lambda}_{p,q}(X_i;Y_i). \tag{10.18}$$

Proof. The proof here is due to Beigi and Nair [16] and is based on applying the one-dimensional BL inequality in (10.15) to each pair of random variables iteratively. To prove (10.17), it suffices to show that if for each $i \in [n]$, there exist a constant $\overline{C}_i$ such that $\langle f_i, g_i\rangle \leq$

$\overline{C}_i\|f_i\|_p\|g_i\|_q$ holds for all nonnegative f_i and g_i defined on $\mathcal{X}$ and $\mathcal{Y}$, then $\langle f,g\rangle \leq \overline{C}\|f\|_p\|g\|_q$ holds for all nonnegative f and g defined on $\mathcal{X}^n$ and $\mathcal{Y}^n$, where $\overline{C} = \prod_{i=1}^n \overline{C}_i$. This point can be shown as follows:

$$\begin{aligned}
\langle f,g\rangle &= \mathbb{E}_{X^{n-1},Y^{n-1}}\left[\mathbb{E}_{X_n,Y_n}[f(X^n)g(Y^n) \mid X^{n-1},Y^{n-1}]\right]\\
&\leq \overline{C}_n\mathbb{E}_{X^{n-1},Y^{n-1}}\left[\|f(X^{n-1},\cdot)\|_p\|g(Y^{n-1},\cdot)\|_q\right]\\
&\leq \overline{C}_n\overline{C}_{n-1}\mathbb{E}_{X^{n-2},Y^{n-2}}\left[\|f(X^{n-2},\cdot)\|_p\|g(Y^{n-2},\cdot)\|_q\right]\\
&\quad\vdots\\
&\leq \overline{C}\|f\|_p\|g\|_q.
\end{aligned}$$

Hence, we have (10.17). The inequality in (10.18) follows similarly. □

Lemma 10.3 (Data processing inequalities). Assume random variables U, X, Y, and V form a Markov chain $U - X - Y - V$ in this order. Then for $p, q \geq 1$,

$$\underline{\Lambda}_{p,q}(X;Y) \leq \underline{\Lambda}_{p,q}(U;V), \tag{10.19}$$

and for $p, q \leq 1$,

$$\overline{\Lambda}_{p,q}(X;Y) \geq \overline{\Lambda}_{p,q}(U;V). \tag{10.20}$$

Moreover, if U and V are deterministic functions of X and Y respectively, then the two inequalities hold for all $p, q \in \mathbb{R}$.

Proof. For any $f : \mathcal{U} \to [0,\infty)$ and $g : \mathcal{V} \to [0,\infty)$, let $\hat{f} : x \in \mathcal{X} \mapsto \mathbb{E}[f(U) \mid X = x]$ and $\hat{g} : y \in \mathcal{Y} \mapsto \mathbb{E}[g(V) \mid Y = y]$. Then we have $\langle f,g\rangle = \langle \hat{f},\hat{g}\rangle$, and by Jensen's inequality, $\|f\|_p \geq \|\hat{f}\|_p$ and $\|g\|_q \geq \|\hat{g}\|_q$ for $p, q \geq 1$, and the directions of these two inequalities are reversed for $p, q \leq 1$. These facts establish (10.19) and (10.20). □

Similarly to the hypercontractivity regions (see Definition 10.4 and Lemma 10.1), the BL exponents also admit rather natural information-theoretic characterizations. Define the function

$$\begin{aligned}
\phi(Q_X,Q_Y) := \inf_{R_X,R_Y}\Big\{&\mathsf{D}(R_X,R_Y\|\pi_{XY}) + \frac{1}{p}D(R_X\|Q_X) - \frac{1}{p}D(R_X\|\pi_X)\\
&+ \frac{1}{q}D(R_Y\|Q_Y) - \frac{1}{q}D(R_Y\|\pi_Y)\Big\},
\end{aligned} \tag{10.21}$$

where according to Convention 10.1, the infimization is taken over all pairs of distributions $(R_X, R_Y) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ such that all the relative entropies in the objective function are finite. Then we have the following information-theoretic characterizations of the forward and reverse BL exponents.

Proposition 10.1. For $p, q \in \mathbb{R}\backslash\{0\}$, if $(X, Y) \sim \pi_{XY}$, then

$$\underline{\Lambda}_{p,q}(X;Y) = \inf_{Q_X,Q_Y} \phi(Q_X, Q_Y) \quad \text{and} \tag{10.22}$$
$$\overline{\Lambda}_{p,q}(X;Y) = \sup_{Q_X,Q_Y} \phi(Q_X, Q_Y).$$

Proof. The proof leverages the following "duality" lemma.

Lemma 10.4 (Duality of Relative Entropy)**.** Let $\{P_i\}_{i=1}^n$ be n probability mass functions on a finite set $\mathcal{X}$. Let $\{s_i\}_{i=1}^n \subset \mathbb{R} \setminus \{0\}$ be nonzero real numbers such that $\sum_{i=1}^n s_i = 1$. Let $c : \mathcal{X} \to \mathbb{R}$ be a function. Define

$$\beta := \sum_{x\in\mathcal{X}} 2^{-c(x)} \left(\prod_{i=1}^n P_i(x)^{s_i} \right).$$

Then we have[2]

$$-\log \beta = \inf_{Q \ll P_i, \forall i\in[n]} \left\{ \sum_{i=1}^n s_i\, D(Q\|P_i) + \mathbb{E}_Q[c(X)] \right\}. \tag{10.23}$$

Moreover, if $0 < \beta < \infty$, the infimization in (10.23) is uniquely attained by the distribution

$$Q^*(x) = \frac{2^{-c(x)}}{\beta} \left(\prod_{i=1}^n P_i(x)^{s_i} \right) \quad \text{for all } x \in \mathcal{X}.$$

This lemma was stated by Shayevitz [154]. It can be proved by using Lagrange multipliers. The generalization of this lemma to arbitrary measurable spaces can be proven by using the nonnegativity of the relative entropy; see [112, Theorem 2.2.3] or [192].

We may assume, by homogeneity, that $\|f\|_p = \|g\|_q = 1$. Without loss of generality, we may also assume, due to Assumption 10.1, that $\mathrm{supp}(f) \subset \mathrm{supp}(\pi_X)$ and $\mathrm{supp}(g) \subset \mathrm{supp}(\pi_Y)$. Hence, we can write

$$f(x)^p = \frac{Q_X(x)}{\pi_X(x)} \quad \text{and} \quad g(y)^q = \frac{Q_Y(y)}{\pi_Y(y)}, \tag{10.24}$$

[2] We adopt the convention $\inf_\emptyset = \infty$, $0 \cdot \infty = 0$, and $0^s = \infty$ for $s < 0$.

for some probability mass functions Q_X and Q_Y. Moreover, since f and g are finite on their supports, Q_X and π_X are mutually absolutely continuous if $p < 0$, and Q_Y and π_Y are mutually absolutely continuous if $q < 0$. From (10.24), we see that

$$\langle f, g \rangle = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \pi_{XY}(x,y) \Big(\frac{Q_X(x)}{\pi_X(x)}\Big)^{1/p} \Big(\frac{Q_Y(y)}{\pi_Y(y)}\Big)^{1/q}. \tag{10.25}$$

Now substituting (10.25) into the definitions of $\underline{\Lambda}_{p,q}$ and $\overline{\Lambda}_{p,q}$, and using Lemma 10.4 (with the identifications $c \leftarrow 0$, $s_1 \leftarrow 1$, $s_2 \leftarrow 1/p$, $s_3 \leftarrow -1/p$, $s_4 \leftarrow 1/q$, $s_5 \leftarrow -1/q$, and $P_1 \leftarrow \pi_{XY}$, $P_2 \leftarrow Q_X\pi_{Y|X}$, $P_3 \leftarrow \pi_{XY}$, $P_4 \leftarrow Q_Y\pi_{X|Y}$, $P_5 \leftarrow \pi_{XY}$), we obtain Proposition 10.1. $\square$

Define the following linear combination of relative entropies

$$\theta(Q_X, Q_Y) := \mathsf{D}(Q_X, Q_Y \| \pi_{XY}) - \frac{1}{p} D(Q_X \| \pi_X) - \frac{1}{q} D(Q_Y \| \pi_Y).$$

By using the tensorization property, the BL exponents also can be written in the following alternative information-theoretic forms in terms of variational characterizations of $\theta(Q_X, Q_Y)$.

Theorem 10.5. For $p, q \in \mathbb{R}\backslash\{0\}$, if $(X, Y) \sim \pi_{XY}$, then

$$\underline{\Lambda}_{p,q}(X;Y) = \begin{cases} \inf\limits_{Q_X,Q_Y} \theta(Q_X, Q_Y) & p, q > 0 \\ -\infty & p < 0 \text{ or } q < 0 \end{cases} \tag{10.26}$$

and

$$\overline{\Lambda}_{p,q}(X;Y) = \begin{cases} \sup\limits_{Q_X,Q_Y} \theta(Q_X, Q_Y) & p, q > 0 \\ \sup\limits_{Q_X} \inf\limits_{Q_Y} \theta(Q_X, Q_Y) & q < 0 < p \\ \sup\limits_{Q_Y} \inf\limits_{Q_X} \theta(Q_X, Q_Y) & p < 0 < q \\ 0 & p, q < 0 \end{cases}. \tag{10.27}$$

For Euclidean spaces, the forward part of this theorem, i.e., (10.26), was derived in Carlen and Cordero-Erausquin [33]. The reverse part of this theorem, i.e., (10.27), for finite alphabets was derived in Beigi and Nair [16] for all $p, q \neq 0$, and also by Liu *et al.* [113] for $p, q > 0$.

The characterizations in (10.26) and (10.27) are consistent with the ones for the hypercontractivity regions given in Theorem 10.1. This can be seen observing that $\underline{\Lambda}_{p,q} \geq 1$ if and only if $(p,q) \in \mathcal{R}_{\mathrm{FH}}(\pi_{XY})$, and $\underline{\Lambda}_{p,q} \leq 1$ if and only if $(p,q) \in \mathcal{R}_{\mathrm{RH}}(\pi_{XY})$. Hence, Theorem 10.1 is indeed a consequence of Theorem 10.5.

Proof of Theorem 10.5. The characterization in (10.26) follows directly from (10.22) by swapping the two infima. We now prove the characterization in (10.27). We first consider the case of $p, q > 0$. On one hand, by setting (R_X, R_Y) in (10.21) to be (Q_X, Q_Y), we have that $\phi(Q_X, Q_Y) \leq \theta(Q_X, Q_Y)$. Hence,

$$\overline{\Lambda}_{p,q}(X;Y) \leq \sup_{Q_X, Q_Y} \theta(Q_X, Q_Y). \tag{10.28}$$

On the other hand, by the tensorization property stated in (10.18) in Lemma 10.2, for $(X^n, Y^n) \sim \pi_{XY}^n$,

$$\begin{aligned}
\overline{\Lambda}_{p,q}(X;Y) &= \frac{1}{n}\overline{\Lambda}_{p,q}(X^n, Y^n) \\
&= \sup_{f,g: \|f\|_p \|g\|_q > 0} -\frac{1}{n} \log \frac{\langle f, g\rangle}{\|f\|_p \|g\|_q} \\
&\geq \max_{\mathcal{A}_n \subset \mathcal{X}^n, \mathcal{B}_n \subset \mathcal{Y}^n} -\frac{1}{n} \log \frac{\pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n)}{\pi_X^n(\mathcal{A}_n)^{1/p} \pi_Y^n(\mathcal{B}_n)^{1/q}},
\end{aligned} \tag{10.29}$$

where in the last line, we restrict f and g to be the indicators of two non-empty sets $\mathcal{A}_n \subset \mathcal{X}^n$ and $\mathcal{B}_n \subset \mathcal{Y}^n$, respectively.

To further lower bound (10.29), we take $(\mathcal{A}_n, \mathcal{B}_n)$ therein to be a pair of type classes $(\mathcal{T}_{T_X^{(n)}}, \mathcal{T}_{T_Y^{(n)}})$ in which the sequence of pairs of types $\{(T_X^{(n)}, T_X^{(n)})\}_{n \in \mathbb{N}}$ converges to some pair of distributions (Q_X, Q_Y) as $n \to \infty$. Then, by Sanov's theorem (see Theorem 1.1),

$$\lim_{n\to\infty} -\frac{1}{n} \log \frac{\pi_{XY}^n\big(\mathcal{T}_{T_X^{(n)}} \times \mathcal{T}_{T_Y^{(n)}}\big)}{\pi_X^n\big(\mathcal{T}_{T_X^{(n)}}\big)^{1/p} \pi_Y^n\big(\mathcal{T}_{T_Y^{(n)}}\big)^{1/q}} = \theta(Q_X, Q_Y). \tag{10.30}$$

Hence, we obtain $\overline{\Lambda}_{p,q}(X;Y) \geq \theta(Q_X, Q_Y)$. Since (Q_X, Q_Y) is arbitrary, we have $\overline{\Lambda}_{p,q}(X;Y) \geq \sup_{Q_X, Q_Y} \theta(Q_X, Q_Y)$. Therefore, (10.27) holds.

We omit the proofs for other cases, since they are similar to the above argument. □

An interesting observation arising from this proof is the following. For $p, q > 0$, by combining (10.28) and (10.29), we obtain

$$\max_{\mathcal{A}_n \subset \mathcal{X}^n, \mathcal{B}_n \subset \mathcal{Y}^n} -\frac{1}{n} \log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) + \frac{1}{np} \log \pi_X^n(\mathcal{A}_n) + \frac{1}{nq} \log \pi_Y^n(\mathcal{B}_n) \leq \sup_{Q_X, Q_Y} \theta(Q_X, Q_Y). \tag{10.31}$$

Furthermore, as shown in (10.30), by appealing to Sanov's theorem, this inequality is *asymptotically tight* (which means that as $n \to \infty$, the limits of the left- and right-hand sides are equal). Similarly, for $p, q > 0$, one can observe that

$$\min_{\mathcal{A}_n \subset \mathcal{X}^n, \mathcal{B}_n \subset \mathcal{Y}^n} -\frac{1}{n} \log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) + \frac{1}{np} \log \pi_X^n(\mathcal{A}_n) + \frac{1}{nq} \log \pi_Y^n(\mathcal{B}_n) \geq \inf_{Q_X, Q_Y} \theta(Q_X, Q_Y), \tag{10.32}$$

and this inequality is also asymptotically tight by Sanov's theorem [49].

As a consequence of (10.31) and (10.32), we find that certain sequences of $\{0, 1\}$-valued functions attain the BL exponents. Hence, a BL inequality holds for all nonnegative functions if and only if any of its multi-dimensional extensions hold for any $\{0, 1\}$-valued functions. In addition to the set of $\{0, 1\}$-valued functions, one can also use the following construction of functions to assert the (asymptotic) optimality of a BL inequality. We can first identify a optimal pair (f^*, g^*) for the one-dimensional case. By the tensorization property of the BL exponents (Theorem 10.2), the n-fold product of (f^*, g^*) also constitutes an optimal pair that allows us to assert the optimality of a BL inequality. In contrast, the asymptotic optimality of $\{0, 1\}$-valued functions is advantageous in our quest to prove the strong SSE theorem (Theorem 8.11) as will be done in Section 10.3.

10.2.3 Single-Function Versions

The BL inequalities discussed in Section 10.2.2 involve *two* nonnegative functions. In the literature, there exist *single-function* versions of BL inequalities and they have been shown to be equivalent to their two-function counterparts (as was discussed in the context of the DSBS in Section 8.3.2). We now introduce the single-function versions of BL inequalities. First recall from (8.55) that the *conditional expectation operator* induced by $\pi_{X|Y}$ is the operator that maps a function $f : \mathcal{X} \to \mathbb{R}$ to the function

$$y \in \mathcal{Y} \mapsto \pi_{X|Y=y}(f) := \mathbb{E}[f(X) \,|\, Y = y] = \sum_{x\in\mathcal{X}} \pi_{X|Y}(x|y) f(x).$$

Then, given a joint distribution π_{XY} and two real numbers p and q, for any nonnegative function $f : \mathcal{X} \to [0, \infty)$, the single-function versions of the BL inequalities read

$$\|\pi_{X|Y}(f)\|_q \le \overline{C}\, \|f\|_p \quad \text{and} \tag{10.33}$$

$$\|\pi_{X|Y}(f)\|_q \ge \underline{C}\, \|f\|_p \tag{10.34}$$

for some constants $\overline{C}$ and $\underline{C}$.

We remark that (10.33) and (10.34) are in fact equivalent to the *strong data processing inequalities* for the Rényi divergence [142]. The latter concerns the tradeoff between $D_p(Q_X\|\pi_X)$ and $D_q(Q_Y\|\pi_Y)$, where Q_Y represents the output distribution induced by the input distribution Q_X and the stochastic kernel $\pi_{Y|X}$, i.e., $Q_X \to \pi_{Y|X} \to Q_Y$. The equivalence follows since we can set $f = Q_X/\pi_X$ and observe that

$$\log \|f\|_p = \frac{1}{p'}\, D_p(Q_X\|\pi_X) \tag{10.35}$$

and

$$\begin{aligned}
\log \|\pi_{X|Y}(f)\|_q &= \frac{1}{q}\, \log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} \frac{Q_X(x)}{\pi_X(x)} \pi_{X|Y}(x|y) \right)^q \pi_Y(y) \\
&= \frac{1}{q}\, \log \sum_{y\in\mathcal{Y}} \left(\frac{Q_Y(y)}{\pi_Y(y)} \right)^q \pi_Y(y) \\
&= \frac{1}{q'}\, D_q(Q_Y\|\pi_Y).
\end{aligned}$$

For more details, see the papers by Raginsky [142] and Yu [192].

The promised equivalence between the single- and two-function versions of the BL inequalities is formalized in the following proposition.

Proposition 10.2. Inequality (10.33) for $q \ge 1$ holds if and only if (10.15) holds but with q in the latter replaced by its Hölder conjugate $q' = \frac{q}{q-1}$. Similarly, inequality (10.34) for $q \le 1$ holds if and only if (10.16) holds but with q in the latter replaced by its Hölder conjugate q'.

Proof. By Hölder's inequality, for any $\hat{g} : \mathcal{Y} \to [0, \infty)$, it holds that

$$\|\hat{g}\|_q = \begin{cases} \sup_{g:\|g\|_{q'}>0} \frac{\langle \hat{g}, g\rangle}{\|g\|_{q'}} & q \geq 1 \\ \inf_{g:\|g\|_{q'}>0} \frac{\langle \hat{g}, g\rangle}{\|g\|_{q'}} & q \leq 1 \end{cases},$$

where $1' = \infty$ and $1' = -\infty$ for the first and second clauses respectively. Setting $\hat{g}$ to be $\pi_{X|Y}(f)$, we obtain the following equivalences: For $q \geq 1$,

$$\sup_{f:\|f\|_p>0} \frac{\|\pi_{X|Y}(f)\|_q}{\|f\|_p} = \sup_{(f,g):\|f\|_p>0,\|g\|_{q'}>0} \frac{\langle f, g\rangle}{\|f\|_p\|g\|_{q'}}, \quad (10.36)$$

and for $q \leq 1$,

$$\inf_{f:\|f\|_p>0} \frac{\|\pi_{X|Y}(f)\|_q}{\|f\|_p} = \inf_{(f,g):\|f\|_p>0,\|g\|_{q'}>0} \frac{\langle f, g\rangle}{\|f\|_p\|g\|_{q'}}. \quad (10.37)$$

By the equivalence in (10.36), for $q \geq 1$, the single-function version of BL inequality in (10.33) is equivalent to the two-function version in (10.15) with $\overline{C}$ and p unchanged but with q replaced by its Hölder conjugate q'. Similarly, for $q \leq 1$, by the equivalence in (10.37), the single-function version of BL inequality in (10.34) is equivalent to the two-function version in (10.16) with $\underline{C}$ and p unchanged but with q replaced by q'. □

10.3 Connections to the NICD Problem and q-Stability

As observed in the proof of Theorem 10.5, certain sequences of $\{0, 1\}$-valued functions attain the BL exponents. We now provide a detailed discussion on this observation. We also discuss the connections between the BL exponents and the NICD problem (Section 8), as well as the q-stability problem (Section 9).

Recall the general version of the strong SSE theorem (Theorem 8.11) and the general version of the strong q-stability theorem (Theorem 9.15). For the LD exponents $\underline{\Upsilon}_{\mathrm{LD}}^{(n)}$ and $\overline{\Upsilon}_{\mathrm{LD}}^{(n)}$ defined in (8.69) and (8.70), the strong SSE theorem states that for π_{XY} defined on a finite alphabet, any $n \geq 1$, $\alpha \in (0, \alpha_{\max}(\pi_X)]$, and $\beta \in (0, \beta_{\max}(\pi_Y)]$, it holds that

$$\underline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \mathbb{L}[\underline{\Upsilon}_{\mathrm{LD}}](\alpha,\beta) \quad \text{and} \tag{10.38}$$

$$\overline{\Upsilon}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \mathbb{U}[\overline{\Upsilon}_{\mathrm{LD}}](\alpha,\beta). \tag{10.39}$$

Moreover, the inequalities in (10.38) and (10.39) are asymptotically tight in the limit as $n \to \infty$. We now provide a proof of the strong SSE theorem by leveraging its connections to the information-theoretic characterizations of BL exponents.

Proof of Theorem 8.11. Observe that (10.31) and (10.32) for $p, q > 0$ can be rewritten as follows. For all $\mathcal{A}_n \subset \mathcal{X}^n$ and $\mathcal{B}_n \subset \mathcal{Y}^n$,

$$\begin{aligned} -\frac{1}{n}\log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) + \frac{1}{np}\log \pi_X^n(\mathcal{A}_n) + \frac{1}{nq}\log \pi_Y^n(\mathcal{B}_n) \\ \geq \inf_{s,t\geq 0} \underline{\varphi}(s,t) - \frac{s}{p} - \frac{t}{q} \geq \inf_{s,t\geq 0} \underline{\Upsilon}(s,t) - \frac{s}{p} - \frac{t}{q}, \end{aligned} \tag{10.40}$$

where $\underline{\varphi}$ and $\underline{\Upsilon}$ are defined in (10.1) and (10.5) respectively. Analogously, for all $\mathcal{A}_n \subset \mathcal{X}^n$ and $\mathcal{B}_n \subset \mathcal{Y}^n$,

$$\begin{aligned} -\frac{1}{n}\log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) + \frac{1}{np}\log \pi_X^n(\mathcal{A}_n) + \frac{1}{nq}\log \pi_Y^n(\mathcal{B}_n) \\ \leq \sup_{s,t\geq 0} \overline{\varphi}(s,t) - \frac{s}{p} - \frac{t}{q} \leq \sup_{s,t\geq 0} \overline{\Upsilon}(s,t) - \frac{s}{p} - \frac{t}{q}, \end{aligned} \tag{10.41}$$

where $\overline{\varphi}$ and $\overline{\Upsilon}$ are defined in (10.2) and (10.6) respectively. For any $(\mathcal{A}_n, \mathcal{B}_n)$, set $a := -\frac{1}{n}\log \pi_X^n(\mathcal{A}_n)$ and $b := -\frac{1}{n}\log \pi_Y^n(\mathcal{B}_n)$. Let (u, v) be a subgradient[3] of $\underline{\Upsilon}$ at (a, b). Since $\underline{\Upsilon}$ is convex and *nondecreasing*, $u, v \geq 0$. Hence, by definition of the subgradient,

$$\inf_{s,t\geq 0} \underline{\Upsilon}(s,t) - us - vt = \underline{\Upsilon}(a,b) - ua - vb. \tag{10.42}$$

Substituting $p = 1/u$ and $q = 1/v$ into (10.40) and utilizing (10.42), we have

$$-\frac{1}{n}\log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) \geq \underline{\Upsilon}(a,b).$$

[3]Let $\mathcal{I} \subset \mathbb{R}^d$ be convex. A vector $\mathbf{g} \in \mathbb{R}^d$ is a *subgradient* of $f : \mathcal{I} \to \mathbb{R}$ at $\mathbf{x} \in \mathcal{I}$ if for all $\mathbf{z} \in \mathcal{I}$, $f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x} \rangle$.

Similarly, by using (10.41), we have

$$-\frac{1}{n}\log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) \le \overline{\Upsilon}(a,b).$$

Hence,

$$\underline{\Upsilon}^{(n)}(\alpha,\beta) \ge \min_{a\ge\alpha, b\ge\beta} \underline{\Upsilon}(a,b) = \underline{\Upsilon}(\alpha,\beta) \quad \text{and}$$
$$\overline{\Upsilon}^{(n)}(\alpha,\beta) \le \max_{a\le\alpha, b\le\beta} \overline{\Upsilon}(a,b) = \overline{\Upsilon}(\alpha,\beta).$$

Finally, the asymptotic tightness of (8.65) and (8.66) can be verified by appealing to Sanov's theorem (Theorem 1.1). □

The strong SSE theorem (Theorem 8.11) can be further strengthened if the *exact* values of the marginal probabilities are given, instead of only bounds as in the definitions of $\underline{\Upsilon}^{(n)}$ and $\overline{\Upsilon}^{(n)}$. It has been shown in [192] that for all $\mathcal{A}_n \subset \mathcal{X}^n$ and $\mathcal{B}_n \subset \mathcal{Y}^n$,

$$\underline{\Upsilon}(a,b) \le -\frac{1}{n}\log \pi_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) \le \mathbb{U}[\overline{\varphi}](a,b), \tag{10.43}$$

where $a = -\frac{1}{n}\log \pi_X^n(\mathcal{A}_n)$, $b = -\frac{1}{n}\log \pi_Y^n(\mathcal{B}_n)$, and $\overline{\varphi}$ was defined in (10.2). Moreover, the lower and upper bounds in (10.43) are asymptotically tight as $n \to \infty$.

A similar relation can be found between the single-function version of the BL exponents and the notion of q-stability discussed in Section 9.1.1. Following steps similar to the proof for the strong SSE theorem, one can prove the strong q-stability theorem (Theorem 9.15). We omit the details here. Furthermore, similarly to the strong SSE theorem, the strong q-stability theorem can be further strengthened if the marginal probabilities are specified. For details, see [192].

10.4 Logarithmic Sobolev Inequalities

We discuss the *logarithmic Sobolev* (or *log-Sobolev*) *inequalities* in this section. It will be seen (from Theorem 10.6) that such inequalities turn out to be *equivalent*, in sense to be made precise, to the hypercontractivity inequalities (cf. Section 10.2.1). We will also focus on information-theoretic characterizations of certain log-Sobolev inequalities. For more details on the classical aspects of this rich topic, the reader is referred to Raginsky and Sason [143] and Ledoux [105]. The results in this section serve as important elements of the proofs of the main results in Section 10.5 in which the classic hypercontractivity inequalities are strengthened. This section thus forms a bridge between the classic hypercontractivity inequalities and their strengthened versions.

10.4.1 Preliminaries on Dirichlet forms and Entropies

Let $\mathcal{X} = \mathcal{Y}$. As assumed, $\mathcal{X}$ is a finite set. Let $\mathbf{L}$ be a $|\mathcal{X}| \times |\mathcal{X}|$ matrix (a linear operator acting on real-valued functions defined on $\mathcal{X}$) such that $L_{x,y} \geq 0$ for $x \neq y$ and $\sum_{y \in \mathcal{X}} L_{x,y} = 0$ for all x. Let $T_t := \mathrm{e}^{t\mathbf{L}}$ $(t \geq 0)$ be a matrix induced by $\mathbf{L}$, where $\mathrm{e}^{\mathbf{A}}$ denotes the *matrix exponential* of $\mathbf{A}$. The operator T_t is known as a *Markov operator*, which is one that sends a real-valued function on $\mathcal{X}$ to another real-valued function on $\mathcal{X}$. In addition, $\{T_t\}_{t\geq 0}$ forms a *Markov semigroup*, since it satisfies the semigroup property, namely that $T_{t+s} = T_t T_s = T_s T_t$ for all $s, t \geq 0$. For more details on Markov operators and Markov semigroups, the reader is referred to Bakry, Gentil, and Ledoux [9] and Rudnicki, Pichór, and Tyran-Kamińska [145].

Let π be a stationary distribution corresponding to $\{T_t\}_{t\geq 0}$, i.e., $\pi = \pi T_t$ for all $t \geq 0$ or, equivalently, $\pi\mathbf{L} = \mathbf{0}$. We can regard π and T_t (for a fixed $t \geq 0$) as corresponding to π_Y and $\pi_{X|Y}$ respectively. As such, the y^{th} row of the matrix T_t is $\pi_{X|Y}(\cdot|y)$. As usual, denote the inner product for two real-valued functions f and g defined on $\mathcal{X}$ as $\langle f, g\rangle_\pi := \mathbb{E}_\pi[fg] = \sum_{x\in\mathcal{X}} \pi(x) f(x) g(x)$.

Definition 10.6. The *Dirichlet form* of $\{T_t\}_{t\geq 0}$ is

$$\mathcal{E}(f, g) := -\sum_{(x,y)\in\mathcal{X}^2} L_{x,y} f(y) g(x) \pi(x) = -\langle \mathbf{L}f, g\rangle_\pi, \tag{10.44}$$

where $(\mathbf{L}f)(x) := \sum_{y\in\mathcal{X}} L_{x,y} f(y)$. The *normalized Dirichlet form* of $\{T_t\}_{t\geq 0}$ is

$$\overline{\mathcal{E}}(f, g) := \frac{\mathcal{E}(f, g)}{\langle f, g\rangle_\pi}.$$

We now extend the definitions of the Dirichlet form and its normalized version to the n-dimensional Cartesian product space $\mathcal{X}^n$. Let $T_t^{\otimes n}$ be the product semigroup on $\mathcal{X}^n$ induced by T_t. Recall from Section 9.3.1, that given a vector $x^n \in \mathcal{X}^n$, let $x^{\setminus k} := (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n) \in \mathcal{X}^{n-1}$ be the subvector of x^n with the k^{th} coordinate removed. For two real-valued functions f and g defined on $\mathcal{X}^n$, let

$$\psi(x^{\setminus k}) := \mathcal{E}\big(f(x^{\setminus k}, \cdot), g(x^{\setminus k}, \cdot)\big)$$

be the action of the Dirichlet form $\mathcal{E}$ on the k^{th} coordinates of f and g with other coordinates held fixed. Then, the Dirichlet form of f and g and its normalized version are respectively given by

$$\begin{aligned}\mathcal{E}_n(f,g) &:= \sum_{k=1}^{n} \sum_{x^{\setminus k} \in \mathcal{X}^{n-1}} \psi(x^{\setminus k}) \prod_{j \in [n] \setminus \{k\}} \pi(x_j) \quad \text{and} \\ \overline{\mathcal{E}}_n(f,g) &:= \frac{\mathcal{E}_n(f,g)}{\langle f, g\rangle_{\pi^n}}.\end{aligned}$$

In addition to the Dirichlet form, the other quantity involved in log-Sobolev inequalities is the *entropy* of a nonnegative function f.

Definition 10.7. For a nonnegative function f, the *entropy* and the *normalized entropy* of f are respectively defined as

$$\mathrm{Ent}(f) := \mathbb{E}_\pi[f \ln f] - \mathbb{E}_\pi[f] \ln \mathbb{E}_\pi[f] \quad \text{and} \quad \overline{\mathrm{Ent}}(f) := \frac{\mathrm{Ent}(f)}{\mathbb{E}_\pi[f]}.$$

Note that these notions of entropy and normalized entropy are commonly encountered in functional analysis; see, for example, Ledoux [105]. They are related to, but not the same as the Shannon entropy in classical information theory. Indeed, they bear more similarity to the relative entropy, in the sense that if f is the Radon–Nikodym derivative $\mathrm{d}Q/\mathrm{d}\pi$ of a distribution Q with respect to π (i.e., the function $x \in \mathcal{X} \mapsto Q(x)/\pi(x)$ for the finite alphabet case), then the entropy (and also the normalized entropy) of f is equal to the relative entropy of Q from π, i.e., $D(Q\|\pi)$. By Jensen's inequality, both the entropy and the normalized entropy are nonnegative.

10.4.2 Log-Sobolev Inequalities and Their Properties

The *log-Sobolev* inequalities quantify the relation between the Dirichlet form of a Markov semigroup for an arbitrary nonnegative function f and a composite function $g = \varphi \circ f$ for some given $\varphi : [0, \infty) \to [0, \infty)$, and the entropy of f. For $p \in \mathbb{R} \setminus \{0, 1\}$, let

$$c_p := \frac{p^2}{4(p-1)}.$$

Following the definitions in Mossel, Oleszkiewicz, and Sen [125], we define log-Sobolev inequalities as follows.

Definition 10.8. For $p \in \mathbb{R}\backslash\{0, 1\}$, the *$p$-log-Sobolev inequality with constant C* is

$$\mathrm{Ent}(f^p) \le C\, c_p\, \mathcal{E}(f, f^{p-1}) \tag{10.45}$$

for nonnegative f if $p > 1$ and for positive f if $p < 1$.For $p = 1$, the 1-*log-Sobolev inequality with constant C* for positive f is

$$\mathrm{Ent}(f) \le \frac{C}{4}\, \mathcal{E}(f, \ln f).$$

For $p = 0$, the 0-*log-Sobolev inequality with constant C* for positive f is

$$\mathrm{Var}(\ln f) \le -\frac{C}{2}\, \mathcal{E}\Big(f, \frac{1}{f}\Big).$$

The cases corresponding to $p = 0$ and $p = 1$ of the p-log-Sobolev inequality are the limiting cases of the p-log-Sobolev inequality for $p \in \mathbb{R}\backslash\{0, 1\}$ with the same constant C.

We now connect the log-Sobolev inequality and the classic hypercontractivity inequalities in (10.33) and (10.34) with $\overline{C}$ and $\underline{C}$ set to 1. Indeed, we will see from Theorem 10.6 that the log-Sobolev inequalities are *differential versions* of the hypercontractivity inequalities evaluated at $t = 0$. This theorem is a classical result due to Gross [69], and various proofs can be found in [6]–[8], [69], [125].

Here we provide a short self-contained proof.

Theorem 10.6 (Differential relationship between log-Sobolev and hypercontractivity inequalities)**.** Let C be a positive constant. Let $q : [0, \infty) \to \mathbb{R}$ be defined as

$$q(t) = 1 + (p-1)\, \mathrm{e}^{4t/C}. \tag{10.46}$$

(a) Fix $p > 1$. If for any $r \in [p, \infty)$, the r-log Sobolev inequality is satisfied with constant C, then for any $t > 0$,

$$\|T_t f\|_{q(t)} \leq \|f\|_p \quad \text{for all } f \geq 0, \tag{10.47}$$

where $(T_t f)(x) = \sum_y T_t(x, y) f(y)$.

(b) Fix $p < 1$. If for any $r \in (-\infty, p]$, the r-log-Sobolev inequality is satisfied with constant C, then for any $t > 0$,

$$\|T_t f\|_{q(t)} \geq \|f\|_p \quad \text{for all } f \geq 0. \tag{10.48}$$

(c) Conversely, if (10.47) holds for $p > 1$ or (10.48) holds for $p < 1$, then the p-log-Sobolev inequality is satisfied with constant C.

The inequalities in (10.47) and (10.48) are respectively equivalent to the fact that[4] $(p, q(t)')$ belongs to the forward and reverse hypercontractivity regions of the joint distributions (Definition 10.4) induced by (T_t, π) for any $t > 0$. In fact, the relations between the BL inequalities and generalized p-log-Sobolev inequalities can also be established. For details, the reader is referred to [6]–[8], [69], [125].

Proof Sketch of Theorem 10.6. We first prove Statement (a) in which we assume that $p > 1$. Define the function $\zeta : [0, \infty)^2 \to \mathbb{R}$ as

$$\zeta(t, s) := \ln \|T_t f\|_{\frac{1}{s}}.$$

Then, one can check by direct differentiation that

$$\frac{\partial \zeta}{\partial s} = -\overline{\mathrm{Ent}}\big((T_t f)^{\frac{1}{s}}\big) \quad \text{and} \quad \frac{\partial \zeta}{\partial t} = -\overline{\mathcal{E}}_n\big(T_t f, (T_t f)^{\frac{1}{s}-1}\big). \tag{10.49}$$

We define $\xi(t) := 1/q(t)$, and hence, $\xi(0) = 1/p$ (refer to (10.46)). Therefore, by (10.49) and the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t, \xi(t)) = -\overline{\mathrm{Ent}}\Big((T_t f)^{\frac{1}{\xi(t)}}\Big)\xi'(t) - \overline{\mathcal{E}}_n\Big(T_t f, (T_t f)^{\frac{1}{\xi(t)}-1}\Big). \tag{10.50}$$

Observe that

$$\xi'(t) = \frac{-4(p-1)\mathrm{e}^{4t/C}}{C\, q(t)^2}.$$

[4] Here $q(t)'$ denotes the the Hölder conjugate of $q(t)$. In contrast, we use $q'(t)$ to denote the derivative of q evaluated at t.

It also holds that $\xi'(t) \geq -1/(C\, c_{q(t)})$ for all $t \geq 0$. Combining this with (10.50) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t,\xi(t)) \leq \overline{\mathrm{Ent}}\big((T_t f)^{\frac{1}{\xi(t)}}\big)\frac{1}{C\, c_{q(t)}} - \overline{\mathcal{E}}_n\big(T_t f, (T_t f)^{\frac{1}{\xi(t)}-1}\big). \quad (10.51)$$

On the other hand, by assumption, for any $r \in [p, \infty)$, the r-log-Sobolev inequality holds, i.e.,

$$\overline{\mathrm{Ent}}(g^r) \leq C\, c_r\, \overline{\mathcal{E}}_n(g, g^{r-1}) \quad \text{for all } g \geq 0. \quad (10.52)$$

Substituting g and r for $T_t f$ and $q(t) = 1/\xi(t)$ respectively into (10.52) and combining the resultant inequality with (10.51) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\zeta(t,\xi(t)) \leq 0 \quad \text{for all } t \geq 0. \quad (10.53)$$

Finally, by integrating both sides of (10.53) from 0 to t, we have

$$\ln \|T_t f\|_{\frac{1}{\xi(t)}} - \ln \|f\|_{\frac{1}{\xi(0)}} \leq 0,$$

which is precisely the hypercontractivity inequality in (10.47).

Statement (b) follows analogously but the directions of the inequalities above are reversed. Statement (c) follows by first differentiating the hypercontractivity inequalities and evaluating them at $t = 0$. Then, we can recover the p-log-Sobolev inequalities from them. □

It is also well-known (see, for example, Ledoux [105]) that the p-log-Sobolev inequality satisfies the *tensorization property*.

Proposition 10.3. If a certain p-log-Sobolev inequality with constant C holds for (T_t, π), then the p-log-Sobolev inequality with the *same* constant holds for the product semigroup $(T_t^{\otimes n}, \pi^n)$.

Proof. We provide an information-theoretic proof for this proposition. We start by characterizing the optimal constant in the p-log-Sobolev inequality in terms of information-theoretic quantities. For the product semigroup $(T_t^{\otimes n}, \pi^n)$, the optimal constant is for $p \in \mathbb{R}\backslash\{0, 1\}$

$$C^*_{p,n} := \sup_{f\,:\, c_p\, \mathcal{E}_n(f, f^{p-1}) > 0} \frac{\mathrm{Ent}(f^p)}{c_p\, \mathcal{E}_n(f, f^{p-1})}.$$

Since (T_t, π) is a special case of $(T_t^{\otimes n}, \pi^n)$ with n set to 1, $C^*_{p,1}$ is the optimal constant for the semigroup (T_t, π).

For a given Q_{X^n}, define the k^{th} "likelihood ratio"

$$\ell_k(y, x^{\setminus k}) := \frac{Q_{X_k|X^{\setminus k}}(y|x^{\setminus k})}{\pi(y)} \quad \text{for all } (y, x^{\setminus k}) \in \mathcal{X} \times \mathcal{X}^{n-1}.$$

If we write $f^p/\mathbb{E}[f^p] = Q_{X^n}/\pi^n$ for a distribution $Q_{X^n} \ll \pi^n$, then

$$\operatorname{Ent}(f^p) = D(Q_{X^n} \| \pi^n) \quad \text{and} \tag{10.54}$$

$$\mathcal{E}_n(f, f^{p-1}) = \sum_{k=1}^n \mathbb{E}_{\pi^{n-1}}[\eta(X^{\setminus k})], \tag{10.55}$$

where

$$\eta(x^{\setminus k}) := -\frac{Q_{X^{\setminus k}}(x^{\setminus k})}{\pi^{n-1}(x^{\setminus k})} \sum_{x,y} L_{x,y} (\ell_k(y, x^{\setminus k}))^{1/p} (\ell_k(x, x^{\setminus k}))^{1/p'} \pi(x).$$

Uniting (10.54) and (10.55), one can obtain the following information-theoretic characterization of $C^*_{p,n}$.

Lemma 10.7. For $n \in \mathbb{N}$ and $p \in \mathbb{R}\setminus\{0, 1\}$, it holds that

$$C^*_{p,n} = \sup_{Q_{X^n} : c_p \sum_{k=1}^n \mathbb{E}_{\pi^{n-1}}[\eta(X^{\setminus k})] > 0} \frac{D(Q_{X^n} \| \pi^n)}{c_p \sum_{k=1}^n \mathbb{E}_{\pi^{n-1}}[\eta(X^{\setminus k})]}. \tag{10.56}$$

Continuing the proof of Proposition 10.3, we notice that, on one hand, by the data processing inequality for the relative entropy, we have

$$\begin{aligned} D(Q_{X^n} \| \pi^n) &= \sum_{k=1}^n D(Q_{X_k|X^{k-1}} \| \pi | Q_{X^{k-1}}) \\ &\le \sum_{k=1}^n D(Q_{X_k|X^{\setminus k}} \| \pi | Q_{X^{\setminus k}}) \\ &= nD(Q_{X_K|X^{\setminus K} K} \| \pi | Q_{X^{\setminus K}|K} Q_K) \\ &= nD(Q_{X|U} \| \pi | Q_U), \end{aligned} \tag{10.57}$$

where $K \sim Q_K := \text{Unif}[n]$ is independent of X^n and $U := (X^{\setminus K}, K)$.

On the other hand, using the definitions of η and ℓ_k, consider,

$$
\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}_{\pi^{n-1}}[\eta(X^{\setminus k})] \\
&= -\sum_{k=1}^{n} \mathbb{E}_{Q_{X^{\setminus k}}}\left[\sum_{x,y} L_{x,y}(\ell_k(y, X^{\setminus k}))^{1/p}(\ell_k(x, X^{\setminus k}))^{1/p'}\pi(x)\right] \\
&= -n\mathbb{E}_{Q_K}\left[\mathbb{E}_{Q_{X^{\setminus K}}}\left[\sum_{x,y} L_{x,y}(\ell_K(y, X^{\setminus K}))^{1/p}(\ell_K(x, X^{\setminus K}))^{1/p'}\pi(x)\middle| K\right]\right] \\
&= n\sum_{u} Q_U(u)\mathcal{E}\left(\left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p}, \left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p'}\right),
\end{aligned}
\tag{10.58}
$$

where the penultimate equality follows from the uniformity of K and the final equality follows from the definition of the Dirichlet form in (10.44) and that of $U = (X^{\setminus K}, K)$. From (10.57) and (10.58), we conclude that the objective function in (10.56) satisfies

$$
\begin{aligned}
&\frac{D(Q_{X^n}\|\pi^n)}{c_p \sum_{k=1}^{n} \mathbb{E}_{\pi^{n-1}}\left[\eta(X^{\setminus k})\right]} \\
&\quad\le \frac{D(Q_{X|U}\|\pi|Q_U)}{c_p \sum_u Q_U(u)\mathcal{E}\left(\left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p}, \left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p'}\right)} \\
&\quad\le \max_u \frac{D(Q_{X|U=u}\|\pi)}{c_p\,\mathcal{E}\left(\left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p}, \left(\frac{Q_{X|U=u}}{\pi}\right)^{1/p'}\right)} && (10.59) \\
&\quad\le C^*_{p,1}, && (10.60)
\end{aligned}
$$

where in (10.59), the maximum is over all u in the alphabet of U (i.e., $\mathcal{X}^{n-1} \times [n]$) such that the denominator is positive, and (10.60) follows from Lemma 10.7 (with n set to 1). Therefore, $C^*_{p,n} \le C^*_{p,1}$ for any n and $p \in \mathbb{R}\backslash\{0,1\}$.

On the other hand, setting Q_{X^n} to be a product distribution Q_X^n in Lemma 10.7, we find that $C^*_{p,n} \ge C^*_{p,1}$ for all $n \in \mathbb{N}$. Combining these two bounds yields $C^*_{p,n} = C^*_{p,1}$ for $p \in \mathbb{R}\backslash\{0,1\}$. By taking limits in p (toward 0 and 1), one deduces that $C^*_{p,n} \ge C^*_{p,1}$ for $p \in \{0,1\}$. Hence, the tensorizaton property holds. □

The information-theoretic method employed in the proof of Proposition 10.3 can be also used to study certain *nonlinear* versions of log-Sobolev inequalities. These inequalities were proposed as a topic for research by Kalai and Linial [91] in 1995. However, there was no progress for over twenty years since the initial proposal of these inequalities until recent works by Samorodnitsky [148], Samorodnitsky [149], and Polyanskiy and Samorodnitsky [139].

Note that (10.45) delineates a certain *linear* relationship between the entropy $\mathrm{Ent}(f^p)$ and the Dirichlet form $\mathcal{E}(f, f^{p-1})$. For $\alpha \geq 0$ and $n \in \mathbb{N}$, let

$$\mathcal{F}_\alpha^{(n)} := \{f : c_p \, \overline{\mathcal{E}}_n(f, f^{p-1}) = n\alpha\}.$$

To study the *nonlinear* tradeoff between the normalized Dirichlet form and the normalized entropy, we define the *log-Sobolev function* as

$$\Xi_p(\alpha) := \sup_{f \in \mathcal{F}_\alpha^{(1)}} \overline{\mathrm{Ent}}(f^p). \tag{10.61}$$

Extending the definition of $\Xi_p(\alpha)$ from T_t to $T_t^{\otimes n}$, we define

$$\Xi_p^{(n)}(\alpha) := \sup_{f \in \mathcal{F}_\alpha^{(n)}} \frac{1}{n} \overline{\mathrm{Ent}}(f^p). \tag{10.62}$$

It would be useful to provide a tight dimension-independent bound for $\Xi_p^{(n)}(\alpha)$. As mentioned in Section 9.4, by *dimension-independent*, we mean that the bound on $\Xi_p^{(n)}(\alpha)$ does not depend on n; in information theory parlance, this is known as a *single-letter* bound. A tight dimension-independent bound for $\Xi_p^{(n)}(\alpha)$ was shown by Polyanskiy and Samorodnitsky [139] in the following theorem.

Theorem 10.8. It holds that for $p \in \mathbb{R} \backslash \{0, 1\}$,

$$\Xi_p^{(n)}(\alpha) \leq \mathbb{U}[\Xi_p](\alpha). \tag{10.63}$$

Moreover, this upper bound is asymptotically tight as $n \to \infty$, which means that

$$\lim_{n \to \infty} \Xi_p^{(n)}(\alpha) = \mathbb{U}[\Xi_p](\alpha). \tag{10.64}$$

If additionally, Ξ_p in (10.61) is concave, then the upper bound in (10.63) is also tight for all finite $n \geq 1$.

This theorem is a strengthening of the (linear) p-log-Sobolev inequality in (10.45) with optimal constant $C_p^* := C_{p,1}^*$ given in (10.56). Here we set n in (10.56) to 1 since the tensorization property holds. Since the function $\mathbb{U}[\Xi_p]$ is nonlinear in general, the inequality in (10.63) is known as the *nonlinear p-log-Sobolev inequality.*

To appreciate the relation between the linear and nonlinear p-log-Sobolev inequalities, one can demonstrate that the optimal constant C_p^* in the (linear) p-log-Sobolev inequality in (10.45) is the right-derivative of $\mathbb{U}[\Xi_p](\alpha)$ at $\alpha = 0$ if $\Xi_p(0) = 0$. If $\Xi_p(0) > 0$, then the linear p-log-Sobolev inequality in (10.45) does not hold for any finite C.

Proof of Theorem 10.8. We follow the same steps as in the proof of Lemma 10.3 up to (10.58). Then, combining these steps with the definition of $\Xi_p^{(n)}(\alpha)$ in (10.62), we find that

$$\begin{aligned}\Xi_p^{(n)}(\alpha) &\le \sup_{Q_{XU}\,:\, c_p\,\mathbb{E}_{Q_U}\left[\mathcal{E}\left((\frac{Q_{X|U}}{\pi})^{1/p},(\frac{Q_{X|U}}{\pi})^{1/p'}\right)\right]=\alpha} D(Q_{X|U}\|\pi_X|Q_U)\\ &= \mathbb{U}[\Xi_p](\alpha).\end{aligned}$$

The asymptotic tightness of (10.64) can be verified by a time-sharing argument (cf. the discussion after Theorem 8.10). □

10.5 Strengthened Hypercontractivity Inequalities

The tools we reviewed in the preceding sections serve as ingredients for the culmination of this section—namely, a strengthened version of the hypercontractivity inequality. For the sake of clarity, we focus on the DSBS. Before doing so, we provide explicit expressions for the linear and nonlinear p-log-Sobolev inequalities particularized to the DSBS.

For the DSBS, $\mathcal{X} = \{0, 1\}$, $\pi = \pi_Y = \mathrm{Bern}(1/2)$ and T_t is the Markov operator induced by $\pi_{X|Y}$ and is given by

$$T_t f(y) = f(y)\frac{1+\mathrm{e}^{-t}}{2} + f(1-y)\frac{1-\mathrm{e}^{-t}}{2} \qquad y \in \{0,1\}. \tag{10.65}$$

Note that the operator $T_t^{\otimes n}$ is the same as T_ρ in Section 8 with $\rho = \mathrm{e}^{-t}$. From (10.65), we know that $L_{x,y} = \mathbb{1}\{x \neq y\} - 1/2$ which is obtained by differentiating T_t with respect to t and evaluating the derivative at $t = 0$. Moreover, the Dirichlet form for this case is

$$\mathcal{E}_n(f,g) = -\frac{1}{2}\langle \Delta f, g\rangle \quad \text{and} \tag{10.66}$$

$$\mathcal{E}_n(f,f) = \frac{2^{-n}}{4} \sum_{(x^n,y^n):x^n\sim y^n} \big(f(x^n) - f(y^n)\big)^2 = \frac{1}{4}\,\mathbf{I}[f]\,, \tag{10.67}$$

where $\Delta f(x^n) := \sum_{y^n:y^n\sim x^n}(f(y^n) - f(x^n))$, and $x^n \sim y^n$ means that $x^n, y^n \in \{0,1\}^n$ differ in exactly one coordinate. Recall that here $\mathbf{I}[f]$ denotes the total influence of f; see Definition 9.9.

For the DSBS, the optimal constant C in the (linear) p-log-Sobolev inequality is 2; see Gross [69]. The (linear) p-log-Sobolev inequality with optimal constant 2 can be derived from the single-function version of the hypercontractivity inequalities for the DSBS given in Theorem 8.5. This can be done by differentiating both sides of the hypercontractivity inequalities with respect to ρ and evaluating the derivative at $\rho = 1$.

10.5.1 Strong Log-Sobolev Inequalities

Polyanskiy and Samorodnitsky [139] proved the dimension-independent *nonlinear* p-log-Sobolev inequalities as stated in the next theorem. Before introducing these inequalities, we define $b_p : [0, \ln 2] \to [0,\infty)$ to be the convex increasing function given by

$$b_p(t) := \begin{cases} \dfrac{\operatorname{sign}(p-1)}{2}\Big(1-y^{\frac{1}{p}}(1-y)^{1-\frac{1}{p}}-y^{1-\frac{1}{p}}(1-y)^{\frac{1}{p}}\Big) & p \neq 0,1 \\ \Big(\dfrac{1}{2}-y\Big)\ln\dfrac{1-y}{y} & p = 1 \end{cases},$$

where $y(t) := h^{-1}(\ln 2 - t)$ and $h^{-1} : [0, \ln 2] \to [0, 1/2]$ is the inverse of the binary entropy function h *with base* e when its domain is restricted to $[0, 1/2]$.

Theorem 10.9 (p-log-Sobolev Inequality for the DSBS). Let $p \in \mathbb{R}\backslash\{0,1\}$. For all $f : \{0,1\}^n \to [0,\infty)$ (and $f > 0$ if $p < 1$), it holds that

$$\frac{1}{n}\operatorname{sign}(p-1)\,\overline{\mathcal{E}}_n(f,f^{p-1}) \geq b_p\Big(\frac{1}{n}\overline{\operatorname{Ent}}(f^p)\Big)\,, \tag{10.68}$$

where the normalized Dirichlet form is given by (10.66). Let $p = 1$. For all $f : \{0,1\}^n \to (0,\infty)$, it holds that

$$\frac{1}{n}\overline{\mathcal{E}}_n(f, \ln f) \geq b_1\Big(\frac{1}{n}\overline{\mathrm{Ent}}(f)\Big), \tag{10.69}$$

The inequality in (10.69) is the limiting case of (10.68) as $p \to 1$. Moreover, these inequalities are sharp in the sense that given p, there exists a nonnegative function f such that the equality holds.

Proof Sketch of Theorem 10.9. Theorem 10.9 follows directly from Theorem 10.8 by observing that b_p is the inverse of Ξ_p when π_{XY} is particularized to the DSBS. Moreover, Ξ_p is concave and increasing. The monotonicity and concavity of Ξ_p (or equivalently, the monotonicity and convexity of b_p) can be shown by calculating the first and second derivatives of Ξ_p (or b_p); see Polyanskiy and Samorodnitsky [139] for more details. Furthermore, the asymptotic sharpness of (10.68) follows directly from the asymptotic sharpness of (10.63). □

Based on Theorem 10.9, we are almost ready to introduce a strengthened version of the forward hypercontractivity inequality shown by Polyanskiy and Samorodnitsky [139]. Before doing so, we would like to discuss an intimate relationship between the linear and nonlinear log-Sobolev inequalities and the edge-isoperimetric inequality given in Theorem 9.5, as promised below (9.38). We first consider the linear log-Sobolev inequality. Consider the DSBS and the case $p = 2$. The linear 2-log-Sobolev inequality with optimal constant $C = 2$ reduces to

$$\mathrm{Ent}(f^2) \leq 2\,\mathcal{E}(f,f) \quad \text{for all } f \geq 0. \tag{10.70}$$

By the tensorization property and utilizing (10.67),

$$\mathrm{Ent}(f^2) \leq 2\,\mathcal{E}_n(f,f) = \frac{1}{2}\,\mathbf{I}[f].$$

Setting f to be a Boolean function with mean a, we obtain

$$\mathbf{I}[f] \geq 2a \ln\Big(\frac{1}{a}\Big). \tag{10.71}$$

Note that in the sharp edge-isoperimetric inequality in (9.38), the logarithm used is log (to the base 2), instead of ln. Hence, in (10.71), there is a multiplicative factor $\log \mathrm{e}$ off from the sharp inequality in (9.38).

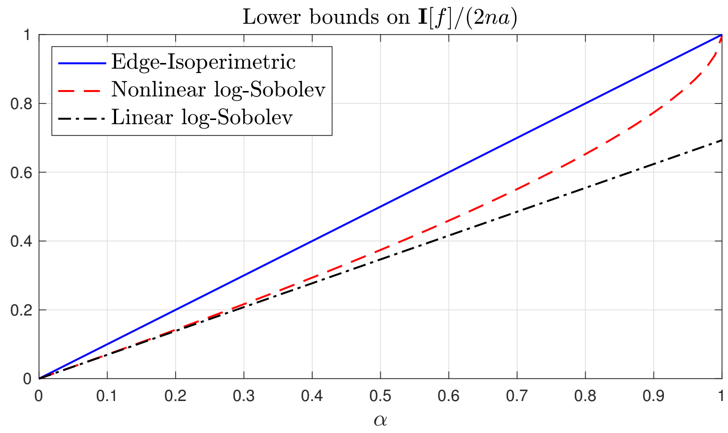

Figure 10.2: Comparison of the distinct parts α (edge-isoperimetric), $\alpha \ln 2$ (linear log-Sobolev) and $2b_2(\alpha)$ (nonlinear log-Sobolev) when a is set to $2^{-\alpha}$.

We next consider the nonlinear log-Sobolev inequality for the DSBS and $p = 2$. For this case, Theorem 10.9 reduces to the statement that for all $f \geq 0$, it holds that

$$\frac{1}{n}\overline{\mathcal{E}}_n(f, f) \geq b_2\Big(\frac{1}{n}\overline{\mathrm{Ent}}(f^2)\Big)\,, \tag{10.72}$$

where $b_2 : [0, \ln 2] \to [0, \infty)$ is the convex increasing function given by

$$b_2(t) = \frac{1 - 2\sqrt{y(1-y)}}{2}$$

with $y = h^{-1}(\ln 2 - t)$ (coinciding with the general definition of b_p). By (10.67) and setting f to be a Boolean function with mean a, we obtain

$$\mathbf{I}[f] \geq 4an\, b_2\Big(\frac{1}{n}\ln\Big(\frac{1}{a}\Big)\Big). \tag{10.73}$$

This inequality is tighter than (10.71), but looser than (9.38). This point can be observed from the facts that b_2 is convex and increasing, $b_2'(0) = 1/2$, $b_2(0) = 0, b_2(\ln 2) = 1/2$, and hence, $t/2 \leq b_2(t) \leq \frac{t}{2\ln 2}$ for all $t \in [0, \ln 2]$. If we consider the case $a = 2^{-\alpha}$, then the bounds in (9.38), (10.71), and (10.73) are $2na\,\alpha$, $4na\, b_2(\alpha)$, and $2na\,\alpha \ln 2$ respectively. Omitting the common factor $2na$, we plot α, $2b_2(\alpha)$, and $\alpha \ln 2$ in Fig. 10.2, which depicts the relations among these three inequalities.

We note that it makes eminent sense that (10.71) and (10.73) are looser than (9.38). This is because that the former two inequalities are derived from the linear and nonlinear log-Sobolev inequalities in (10.70) and (10.72) which are valid not only for Boolean functions, but for *any nonnegative functions*. In contrast, the edge-isoperimetric inequality in (9.38), which is derived by a combinatorial method, is specific to and sharp for Boolean functions.

10.5.2 Strengthened Version of Hypercontractivity Inequalities

We now introduce a strengthened version of forward hypercontractivity inequality due to Polyanskiy and Samorodnitsky [139]. We first introduce an additional definition. For a nonnegative function $f : \mathcal{X}^n \to [0, \infty)$, the *$p$-entropy* of f [192] is defined as

$$\overline{\mathrm{Ent}}_p(f) := \frac{p}{p-1} \log \frac{\|f\|_p}{\|f\|_1}.$$

In fact, if f is the Radon–Nikodym derivative of Q with respect to π, i.e., $f = \mathrm{d}Q/\mathrm{d}\pi$, then $\overline{\mathrm{Ent}}_p(f) = D_p(Q\|\pi)$; also see (10.35). Basic properties of the p-entropy, such as its continuity and monotonicity, can be found in Yu [192]. Let $g : [0, \ln 2] \to [2, 2/\ln 2]$ be defined as

$$g(t) := \frac{2 - 4\sqrt{y(1-y)}}{\ln 2 - h(y)},$$

where $y = h^{-1}(\ln 2 - t)$.

Theorem 10.10. Fix two numbers $1 < p < \infty$ and $0 \le \alpha \le \ln 2$. Then the differential equation in $u : [0, \infty) \to \mathbb{R}$

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = g\Big(\frac{\alpha}{p'}(1 + \mathrm{e}^{-u(t)})\Big)$$

with initial solution $u(0) = \ln(p-1)$ has a unique solution on $[0, \infty)$. Furthermore, for any $f : \{0,1\}^n \to [0, \infty)$ with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \ge \alpha$, we have

$$\|T_t^{\otimes n} f\|_{q(t)} \le \|f\|_p \quad \text{where} \quad q(t) = 1 + \mathrm{e}^{u(t)}\,. \tag{10.74}$$

The core idea of the proof of Theorem 10.10 is to integrate both sides of the nonlinear p-log-Sobolev inequality in Theorem 10.9. It is similar to the proof of Theorem 10.6, and hence, omitted. Theorem 10.10 was used by Ordentlich, Polyanskiy, and Shayevitz [133] to prove the limiting cases as $\rho \downarrow 0$ and $\rho \uparrow 1$ of the NICD problem in the LD regime.

As remarked by Polyanskiy and Samorodnitsky [139], the function g is a smooth, convex, and strictly increasing bijection. Consequently, the function $q(t)$ in (10.74) is smooth and satisfies $q(t) > 1 + (p-1)e^{2t}$ for all $t > 0$. Note that the maximum (and hence best possible) parameter $q(t)$ for the classic forward hypercontractivity is equal to $1 + (p-1)e^{2t}$; see (8.56). Hence, the inequality in (10.74) strictly improves the classic forward hypercontractivity inequality in (8.56). Furthermore, $q(t)$ also satisfies

$$q(t) = p + q'(0)t + \frac{1}{2}q''(0)t^2 + o(t^2), \qquad \text{as } t \to 0\,,$$

where

$$\begin{aligned} q'(0) &= (p-1)g(\alpha), \quad \text{and} \\ q''(0) &= (p-1)\Big(g(\alpha)^2 - g'(\alpha)g(\alpha)\frac{\alpha}{p}\Big). \end{aligned}$$

Since the strengthened version of the hypercontractivity inequality in (10.74) is obtained by integrating both sides of the *sharp* nonlinear p-log-Sobolev inequality in Theorem 10.9, one can observe that (10.74) is locally sharp at $t = 0$ in the following sense. For every $\hat{q}(t)$ such that $\hat{q}(0) = p$ and $\hat{q}'(0) > q'(0)$ there exists a function f with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \geq \alpha$ such that $\|T_t^{\otimes n} f\|_{\hat{q}(t)} > \|f\|_p$ holds for any sufficiently small t. However, the inequality in (10.74) does not appear to be *globally* asymptotically sharp in the sense that there exists a function $\hat{q} : [0, \infty) \to \mathbb{R}$ such that $\hat{q}(t) > q(t)$ for all $t > 0$, and $\|T_t^{\otimes n} f\|_{\hat{q}(t)} \leq \|f\|_p$ holds for all $f : \{0,1\}^n \to [0, \infty)$ with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \geq \alpha$. Recently, a globally sharp inequality was derived by the first author of this monograph [192], who showed that given q, the minimum p such that the inequality $\|T_t^{\otimes n} f\|_q \leq \|f\|_p$ holds for any $f : \{0,1\}^n \to [0, \infty)$ with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \geq \alpha$ is $\alpha/\varphi_q(\alpha)$ (where φ_q is defined in Definition 10.3), in which $\rho = e^{-t}$. Moreover, this sharp bound is asymptotically attained by indicators of Hamming spheres. We compare the sharp bound given by Yu [192] and the bound in Theorem 10.10 in Fig. 10.3. This figure indicates that (10.74) is close to optimal as $p \downarrow 1$.

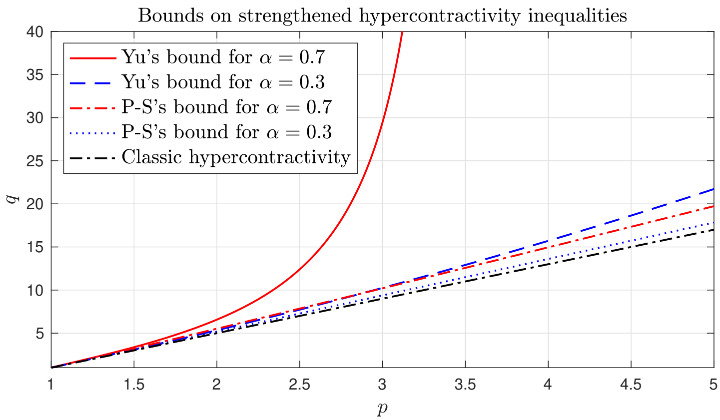

Figure 10.3: Comparison of the sharp bound derived by Yu [192] and the bound in Theorem 10.10 derived by Polyanskiy and Samorodnitsky [139] (P-S), where $\rho = 0.5$. Note that we use q and p to denote $q(t)$ and p respectively in Theorem 10.9.

Along the same lines, it is natural to investigate sharper versions of BL inequalities. Indeed, Polyanskiy posed a conjecture concerning the asymptotically sharp BL inequalities in 2016. This conjecture is stated in Kirshner and Samorodnitsky [97] and reproduced here.

Conjecture 10.1. Fix $\rho \in (0,1)$, $q > 1$, and a scalar $\alpha \in (0, \ln 2)$. Then, there exists

$$p_0 < 1 + \rho^2(q-1)$$

such that for any $p \geq p_0$ the maximum of $\frac{1}{n} \ln \frac{\|T_\rho f\|_q}{\|f\|_p}$ over all nonnegative functions f with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \geq \alpha$ is asymptotically attained by a sequence of functions that are indicators of Hamming spheres with radii converging to some constant as $n \to \infty$.

This conjecture was confirmed in the affirmative by Kirshner and Samorodnitsky [97] for the case $q = 2$. As stated in [97], Conjecture 10.1 for all $q > 1$ was proved by Polyanskiy in an unpublished work [137]. It was also proven independently by the first author of this monograph in [192], [193]. In particular, it was shown that Conjecture 10.1 holds even for $p = 1$. Indeed, the asymptotically sharp bound on $q(t)$ such that $\|T_t^{\otimes n} f\|_{q(t)} \leq \|f\|_p$ holds for any $f : \{0,1\}^n \to [0, \infty)$ with $\frac{1}{n}\overline{\mathrm{Ent}}_p(f) \geq \alpha$ can be characterized by the strong BL inequality derived in the same references. Readers may refer to [192], [193] for the details.

11

Open Problems

We have taken a whirlwind tour of classic and contemporary notions related to the common information between two random variables. In this final section, we list some open problems that represent fertile grounds for future research.

11.1 Open Problems Related to Wyner's Common Information

We now introduce two open problems related to extensions of Wyner's common information.

11.1.1 Rényi Common Information for all Orders

As shown in Part II, the (unnormalized and normalized) Rényi common information forms a bridge between Wyner's common information and the exact common information (see Fig. 5.1). The latter two quantities correspond to the Rényi common information of order 1 (normalized) and order ∞ (unnormalized) respectively. Hence, the Rényi common information of order $\alpha \in [0, \infty]$ is a natural generalization of these quantities. However, the complete characterization of Rényi common information of order α remains open for a large range of α and sources.

Here by "complete characterization", we refer to providing "single-letter expressions". In Theorem 4.3, we present upper and lower bounds on the Rényi common information for $\alpha \in [0, 2] \cup \{\infty\}$. For $\alpha \in (0, 1]$, the unnormalized and normalized Rényi common information of order α are both shown to be equal to Wyner's common information, and hence, it has been completely characterized. However, for $\alpha \in (1, \infty]$, the upper and lower bounds given in Theorem 4.3 only coincide for some special cases, e.g., the case for the DSBS and $\alpha = \infty$, and the case of sources with Wyner-product distributions (cf. Definition 4.3). The complete characterization of Rényi common information for all discrete and continuous sources and for all orders $\alpha \in (1, \infty]$ is a major open problem on this topic. An interesting special case of this open problem is Conjecture 5.1, which concerns the determination of the exact common information (or the unnormalized Rényi common information of order ∞) for jointly Gaussian sources.

11.1.2 Exact Rényi Common Information for all Orders

Another interesting observation from Part II is that the *exact* Rényi common information of order α (originally defined in (7.15))

$$T_{\mathrm{Ex}}^{(\alpha)}(\pi_{XY}) := \lim_{n\to\infty} \frac{G_\alpha(\pi_{XY}^n)}{n}. \tag{11.1}$$

connects the exact common information and the nonnegative rank of a matrix; see Corollary 7.2 and Fig. 7.1. Specifically, the exact common information corresponds to the exact Rényi common information of order 1. Given a bivariate source π_{XY}, if we write its distribution as a matrix $\mathbf{M}$, then the asymptotic exponent of the nonnegative rank of $\mathbf{M}^{\otimes n}$ is the exact Rényi common information of order 0. Hence, the exact Rényi common information of order α in (11.1) simultaneously generalizes both the concepts of the exact common information and the nonnegative rank. This inspires us to define the *nonnegative α-rank*

$$\mathrm{rank}_+^{(\alpha)}(\mathbf{M}) := 2^{G_\alpha(\pi_{XY})}$$

in (7.16) in Section 7.4. This notion extends the concept of common information beyond the realm of information theory. The complete characterization of the exact Rényi common information of order α

remains open. In Corollary 7.2, we provide a single-letter expression for the exact Rényi common information only for the order ∞. Since the exact common information for the DSBS has been completely characterized, the exact Rényi common information of order 1 for the DSBS is completely characterized as well. The complete characterization of the exact Rényi common information of orders $\alpha \in [0, \infty) \setminus \{1\}$ for the DSBS and for $\alpha \in [0, \infty)$ for other sources remains open.

11.2 Open Problems Related to Gács–Körner–Witsenhausen's Common Information

In this section, we introduce several interesting open problems on the extensions of GKW's common information. These extensions mainly concern the q-stability as discussed in Section 9. Recall from Section 9.1.1 that the *noise stability* for a Boolean function $f : \mathcal{X}^n \to \{0, 1\}$ with respect to ρ is

$$\mathbf{S}_\rho[f] = \mathbb{E}[f(X^n)f(Y^n)],$$

where (X^n, Y^n) is a source sequence generated by a DSBS with correlation coefficient $\rho \in [0, 1]$. This concept can be extended to real-valued functions $f : \mathcal{X}^n \to \mathbb{R}$. For the Gaussian source with correlation coefficient $\rho \in [0, 1]$, the noise stability of f can be defined similarly. When there is no ambiguity, for Gaussian sources, we also denote the noise stability of a real-valued function f as $\mathbf{S}_\rho[f]$. For both the DSBS and the Gaussian source, the noise stability and the q-stability satisfy the relation

$$\mathbf{S}_{\rho^2}[f] = \mathbf{S}_\rho^{(2)}[f],$$

for any f and $\rho \in (0, 1)$. The same equation for the DSBS and Boolean functions is given in (9.5).

We classify open problems related to GKW's common information into three sets according to the underlying sources, namely, the DSBS, the Gaussian source, and the so-called ball- and sphere-noise source.

11.2.1 The Doubly Symmetric Binary Source

We introduce four open problems concerning the DSBS.

Determination of $q_{\min}, q_{\max}, \breve{q}_{\min}$ and $\breve{q}_{\max}$

One of main open problems on the q-stability is the determination of the thresholds $q_{\min}$ and $q_{\max}$ for the asymmetric max q-stability and $\breve{q}_{\min}$ and $\breve{q}_{\max}$ for the symmetric max q-stability given in Lemma 9.3 due to Barnes and Özgür [11]. Weaker versions of this open problem are stated in Conjectures 9.1 and 9.2, namely, the symmetric and asymmetric versions of the Mossel–O'Donnell, the Courtade–Kumar, and the Li–Médard conjectures. Although these conjectures for certain ranges of (q, ρ) have been resolved as discussed in Section 9.4, other cases remain wide open. These conjectures are significant since they connect several different fields including discrete Fourier analysis, information theory, discrete probability, etc. Among these conjectures, the Courtade–Kumar conjecture is regarded as one of the most fundamental conjectures at the interface of information theory and the analysis of Boolean functions.

Optimality of Majorities

The general open problem as discussed above on the q-stability appears to be intractable with the current set of analytical tools. A possible strategy to make some progress is to first find the structure of the optimal solutions attaining the max q-stability, and according to this observation, to prove that the optimal solutions belong to a small class of functions. If functions in this class are well-behaved, it is then relatively easy to deduce which Boolean functions *in this class* maximize the q-stability. It has been observed that for odd dimensions n and mean $a = 1/2$, the family of majority functions (defined in Section 9.1.1), which is well-behaved, may be a plausible candidate, since both dictator functions and indicators of Hamming balls are majority functions. Hence, it was conjectured by Mossel and O'Donnell [122] that Maj_n minimizes the symmetric q-stability over all *anti-symmetric* Boolean functions. We state this formally as follows.

Conjecture 11.1 (Optimality of majorities)**.** Consider the DSBS with correlation coefficient $\rho \in (0, 1)$. Fix $q > 1$ and n odd. Then, $\breve{\mathbf{S}}_{\rho}^{(q)}[f]$ is maximized among anti-symmetric Boolean functions f by a majority function Maj_m for some odd number $m \in [n]$.

In the original conjecture [122], q was restricted to be a positive integer. Conjecture 11.1 is weaker than what we hope to resolve the max q-stability problem, since only *anti-symmetric* Boolean functions are considered in this conjecture, instead of all balanced Boolean functions. Indeed, one can consider a more general question whether $\breve{\mathbf{S}}_\rho^{(q)}[f]$ is maximized by a majority function Maj_m among all *balanced* Boolean functions. If the answer is affirmative, it would have significant implications in addressing the max q-stability problem in the sense that it allow us to focus our attention *only on majority functions*.

Stability of Majorities under Bounds on Coefficients

It is also interesting to investigate the noise stability for a specific class of Boolean functions, e.g., the class of functions whose influences or Fourier coefficients are constrained. It is well known that for the majority function Maj_n, all of its Fourier coefficients vanish as $n \to \infty$; see, e.g., [131]. On the other hand, the noise stability $\mathbf{S}_\rho[\mathrm{Maj}_n]$ of Maj_n for the DSBS with correlation coefficient $\rho \in (0, 1)$ satisfies

$$\lim_{n\to\infty} \mathbf{S}_\rho[\mathrm{Maj}_n] = \frac{1}{4} + \frac{\arcsin\rho}{2\pi}. \tag{11.2}$$

This can be shown similarly to (8.18) and (8.73) in which a and b are set to 1/2 and one uses the central limit theorem to approximate the DSBS by a jointly Gaussian source. It has been conjectured by Mossel, O'Donnell, and Oleszkiewicz [123] that for all balanced Boolean functions f with small Fourier coefficients, the noise stability $\mathbf{S}_\rho[f]$ cannot exceed the right-hand side of (11.2) "by too much". This is quantified in the following conjecture.

Conjecture 11.2 (Majority is most stable under bounds on the Fourier coefficients)**.** Consider the DSBS π_{XY} with correlation coefficient $\rho \in (0, 1)$ and let $(X^n, Y^n) \sim \pi_{XY}^n$. Let $f : \{0, 1\}^n \to \{0, 1\}$ be an arbitrary balanced function, i.e., $\mathbb{E}[f] = 2^{-n} \sum_{x^n\in\{0,1\}^n} f(x^n) = 1/2$. Then,

$$\mathbf{S}_\rho[f] \le \frac{1}{4} + \frac{\arcsin\rho}{2\pi} + \varepsilon_\rho\Big(\max_{S\subset[n]} |\hat{f}_S|\Big),$$

where $\varepsilon_\rho(\delta) \downarrow 0$ as $\delta \downarrow 0$ for each fixed $\rho \in (0, 1)$.

A weaker version of this conjecture in which $\max_{S\subset[n]}|\hat{f}_S|$ is replaced by the maximal influence $\max_{i\in[n]}\mathbf{I}_i[f]$ was resolved in [123].

Extracting a Constant or Sublinear Number of Bits

In GKW's common information, the number of bits that is required to be extracted from a source $(X,Y)\sim\pi_{XY}$ is *linear* in the dimension or blocklength n. In contrast, in the Non-Interactive Correlation Distillation (NICD) or the max q-stability problem, only a *single* or a *pair* of random bits is to be extracted. A natural generalization of these two problems is to consider the "intermediate regime" in which the number of bits that one hopes to extract is more than one but sublinear in n. For example, one may wish to extract (a constant) $\ell\geq 2$ bits by using a function $f:\{0,1\}^n\to\{0,1\}^\ell$ applied to X^n and Y^n where (X^n,Y^n) is a source sequence generated by a DSBS. We require f to be balanced (i.e., $f(X^n)\sim\mathrm{Unif}\{0,1\}^\ell$), and at the same time, we aim to maximize the *agreement probability*

$$\Pr\left(f(X^n)=f(Y^n)\right)=\sum_{u^\ell\in\{0,1\}^\ell}\pi_{XY}^n(\mathcal{A}_{u^\ell}\times\mathcal{A}_{u^\ell}),\tag{11.3}$$

where $\mathcal{A}_{u^\ell}=f^{-1}(u^\ell)=\{x^n\in\{0,1\}^n:f(x^n)=u^\ell\}$ for $u^\ell\in\{0,1\}^\ell$ and π_{XY} is the distribution of the DSBS with correlation coefficient $\rho\in(0,1)$. In other words, we wish to find a partition $\{\mathcal{A}_{u^\ell}\}_{u^\ell\in\{0,1\}^\ell}$ of $\{0,1\}^n$ such that each subset $\mathcal{A}_{u^\ell}$ has the same cardinality and (11.3) is maximized. If we naïvely output the first ℓ bits x^ℓ by using the function $f(x^n)=x^\ell$—an indicator of an $(n-\ell)$-subcube (cf. Section 8.2.1)—then the induced agreement probability is exactly $\left(\frac{1+\rho}{2}\right)^\ell$. Indeed, as a consequence of our solution to (the forward part of) Mossel's mean-1/4 stability problem given in Section 8.3.3, for $\ell=2$, the function f that outputs the first two bits attains the maximum of the agreement probability for this problem. In addition, by using the hypercontractivity inequalities in Theorem 8.5, Bogdanov and Mossel [24] showed that the maximal agreement probability

$$\max_{f:\{0,1\}^n\to\{0,1\}^\ell}\Pr\left(f(X^n)=f(Y^n)\right)\leq 2^{-\left(\frac{1-\rho}{1+\rho}\right)\ell}.$$

This upper bound is asymptotically tight as $\ell \to \infty$ in the sense that there exists $f : \{0,1\}^n \to \{0,1\}^\ell$ such that

$$\Pr\left(f(X^n) = f(Y^n)\right) \geq 0.003\sqrt{\frac{2}{(1-\rho)\,\ell}}\, 2^{-\left(\frac{1-\rho}{1+\rho}\right)\ell}.$$

Thus, the exponents of the lower and upper bounds coincide and are equal to $\frac{1-\rho}{1+\rho}$. Determining the *exact* value of the maximum agreement probability over all balanced $\{0,1\}^\ell$-valued functions with fixed $\ell \geq 3$ remains open.

11.2.2 Gaussian Sources

We next introduce two open problems for bivariate Gaussian sources.

Standard Simplex Conjecture

We now consider a Gaussian version of the noise stability problem for balanced $[m]$-valued functions. This problem is analogous to its DSBS counterpart for $\{0,1\}^\ell$-valued functions. In the Gaussian version, we extract a pair of random variables $U = f(X^n)$ and $V = f(Y^n)$ by using a deterministic (measurable) map $f : \mathbb{R}^n \to [m]$ from a pair of length-n vectors (X^n, Y^n) drawn from a bivariate Gaussian source with correlation coefficient $\rho \in (0,1)$. We require f to be balanced in the sense that U, or equivalently V, is uniformly distributed on $[m]$. We aim to maximize the *agreement probability*

$$\Pr(U = V) = \sum_{i=1}^{m} \pi_{XY}^n(\mathcal{A}_i \times \mathcal{A}_i) \tag{11.4}$$

with $\mathcal{A}_i = f^{-1}(i)$ for $i \in [m]$. For this $[m]$-valued function version of Gaussian NICD problem, Isaksson and Mossel [87] posed the *standard simplex conjecture*. Before stating it, we have to introduce some terminology.

A *flat* or *simplex partition* $\{\mathcal{A}_i\}_{i=1}^m$ of $\mathbb{R}^n$ is one in which there exists vectors $\mathbf{a}_0 \in \mathbb{R}^n$ and $\{\mathbf{a}_i\}_{i=1}^m \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

- for all $i, j \in [m]$ such that $i \neq j$, $\mathbf{a}_i$ is not a positive multiple of $\mathbf{a}_j$;

- for all $i \in [m]$,

$$\mathcal{A}_i = \mathbf{a}_0 + \left\{ \mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}_i, \mathbf{x} \rangle = \max_{j \in [m]} \langle \mathbf{a}_j, \mathbf{x} \rangle \right\}.$$

A *standard simplex partition* is a flat partition $\{\mathcal{A}_i\}_{i=1}^m$ where $\|\mathbf{a}_i\|_2 = 1$ for all i and $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = -\frac{1}{m-1}$ for all $i \neq j$.

Conjecture 11.3 (Standard simplex conjecture)**.** Consider the bivariate Gaussian source $(X, Y) \sim \pi_{XY}$ with correlation coefficient $0 < \rho < 1$. Then, among all partitions $\{\mathcal{A}_i\}_{i=1}^m$ of $\mathbb{R}^n$ into $3 \leq m \leq n+1$ measurable parts of equal π_{XY}^n-probability (i.e., $\pi_{XY}^n(\mathcal{A}_i) = 1/m$ for all $i \in [m]$), the agreement probability in (11.4) is *maximized* by standard simplex partitions. Furthermore, for $-1 < \rho < 0$, standard simplex partitions *minimize* the agreement probability in (11.4).

This conjecture was confirmed positively by Heilman [82] for the case $m = 3$ and in the low correlation (i.e., small ρ) regime. Specifically, for $m = 3$ and $n \geq 2$, there exists a function $\rho_0(n) > 0$ such that the conjecture holds for $0 < \rho < \rho_0(n)$. However, for other cases, Conjecture 11.3 remains open.

Symmetric Gaussian Problem

Recall that in the NICD and the max q-stability problem for the Gaussian case with mean $a = 1/2$ (cf. Sections 8.6.2 and 9.6.2), indicators of parallel halfspaces attain the forward joint probability and the max q-stability. Indicators of halfspaces are *anti-symmetric* (or *odd*) in the sense that $f(x^n) = 1 - f(-x^n)$ for almost all $x^n \in \mathbb{R}^n$; see the analogous definition for functions defined on $\{0, 1\}^n$ in Section 9.1.1. It is interesting to ask which *symmetric* (or *even*) functions i.e., those that satisfy $f(x^n) = f(-x^n)$ for almost all $x^n \in \mathbb{R}^n$, maximize the joint probability in the NICD problem or the q-stability in the max q-stability problem. For Gaussian sources, Chakrabarti and Regev [34] posed the following problem.

Problem 11.1 (Symmetric Gaussian problem)**.** Fix $0 < \rho, a, b < 1$. Let $\mathcal{A} \subset \mathbb{R}^n$ and $\mathcal{B} \subset \mathbb{R}^n$ have Gaussian measures a and b, respectively. Furthermore, suppose $\mathcal{A}$ is *centrally symmetric*, i.e., $\mathcal{A} = -\mathcal{A}$. What is the maximum possible value of $\Pr(X^n \in \mathcal{A}, Y^n \in \mathcal{B})$, where X^n and Y^n are ρ-correlated n-dimensional standard Gaussian vectors?

Even though the problem statement requires that $\mathcal{A}$ is centrally symmetric, this problem is equivalent to requiring that *both* $\mathcal{A}$ and $\mathcal{B}$ are centrally symmetric [58]. Indeed, given a set $\mathcal{A}$, the optimal $\mathcal{B}$ that maximizes $\pi_{XY}(\mathcal{A} \times \mathcal{B})$ under the constraint $\pi_Y(\mathcal{B}) = b$ is the set of y such that $\frac{\mathrm{d}\pi_{Y|X}(y|\mathcal{A})}{\mathrm{d}\pi_Y(y)} \geq \lambda$ for some $\lambda > 0$. Hence, if $\mathcal{A}$ is centrally symmetric, so is the optimal $\mathcal{B}$ since for this case,

$$\frac{\mathrm{d}\pi_{Y|X}(y|\mathcal{A})}{\mathrm{d}\pi_Y(y)} = \frac{\mathrm{d}\pi_{Y|X}(-y|\mathcal{A})}{\mathrm{d}\pi_Y(-y)}.$$

It was conjectured in Chakrabarti and Regev [34] and O'Donnell [130] that $\Pr(X^n \in \mathcal{A}, Y^n \in \mathcal{B})$ is maximized by $(\mathbb{B}_r, \mathbb{B}_s)$ or by $(\mathbb{B}_r^{\mathrm{c}}, \mathbb{B}_s^{\mathrm{c}})$ for some appropriate $r, s > 0$, where $\mathbb{B}_r = \{x^n \in \mathbb{R}^n : \|x^n\|_2 \leq r\}$ denotes the ball centered at the origin with radius r. This conjecture was, however, disproved by Heilman [83] in dimensions two and higher.

11.2.3 Ball- and Sphere-Noise Sources

Up to this point, only *memoryless* sources or, equivalently, *product* distributions have been discussed. Extending the NICD and q-stability problems to sources with memory is a more challenging but fruitful endeavor, which may provide unique insights. For simplicity, here we consider *ball-* and *sphere-noise sources* since they behave similarly to memoryless sources in some sense. Hence, results on memoryless sources can be applied to ball- and sphere-noise sources.

NICD for Ball- and Sphere-Noise Sources

We first consider the ball-noise stability problem. Let $\pi_{X^nY^n}$ be a joint distribution on $\{0,1\}^n \times \{0,1\}^n$ such that $\pi_{X^n} = \mathrm{Unif}\{0,1\}^n$ and

$$Y^n = X^n \oplus Z^n = (X_i \oplus Z_i)_{i\in[n]},$$

where $Z^n \sim \mathrm{Unif}(\mathbb{B}_r)$ is independent of X^n and $\oplus$ denotes the modulo-2 sum. Here, $\mathbb{B}_r$ is the Hamming ball centered at 0^n with radius r (cf. Section 8.2.2). The distribution $\pi_{X^nY^n}$ is no longer of a product form since the coordinates of (X^n, Y^n) are correlated through the entries of Z^n. For (n, r, M) such that $1 \le r \le n$ and $1 \le M \le 2^n$, define the *forward joint probability for* $\pi_{X^nY^n}$ (or *maximal ball-noise stability*) as

$$\Gamma^{(n)}_{\mathrm{Ball}}(M, r) := \max_{\substack{f:\{0,1\}^n \to \{0,1\} \\ \Pr(f(X^n)=1)=a}} \Pr(f(X^n) = f(Y^n) = 1),$$

where $(X^n, Y^n) \sim \pi_{X^nY^n}$ and $a = M/2^n$. For fixed $a, \beta \in (0, 1)$, define their upper and lower limits for *even* radii as $n \to \infty$ as

$$\overline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,even}}(a, \beta) := \limsup_{n\to\infty} \Gamma^{(n)}_{\mathrm{Ball}}\left(\lfloor a2^n \rfloor, 2\left\lfloor \frac{\beta n}{2} \right\rfloor\right) \quad \text{and} \tag{11.5}$$

$$\underline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,even}}(a, \beta) := \liminf_{n\to\infty} \Gamma^{(n)}_{\mathrm{Ball}}\left(\lfloor a2^n \rfloor, 2\left\lfloor \frac{\beta n}{2} \right\rfloor\right). \tag{11.6}$$

The limits for *odd* radii, denoted by $\overline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,odd}}$ and $\underline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,odd}}$, can be defined analogously but with $2\lfloor \frac{\beta n}{2} \rfloor$ in (11.5) and (11.6) replaced by $2\lfloor \frac{\beta n}{2} \rfloor + 1$. In many information-theoretic problems (e.g., error exponents for channel coding), when the dimension n is sufficiently large, the uniform distribution on the Hamming ball $\mathbb{B}_r$ can be thought of as the n-fold product of the Bernoulli distribution $\mathrm{Bern}(r/n)$. This inspires the first author of this monograph to pose the following conjecture in [194].

Conjecture 11.4 (NICD for ball-noise sources). For $a, \beta \in (0, 1/2)$,

$$\underline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,even}}(a, \beta) = \overline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,even}}(a, \beta) = \overline{\Gamma}^{(\infty)}(a, a), \tag{11.7}$$

where $\overline{\Gamma}^{(\infty)}$ is the asymptotic forward joint probability for the DSBS with correlation coefficient $\rho = 1 - 2\beta$; see its definition in (8.6).

Conjecture 11.4 pertains only to even radii. For odd radii, Yu [194] showed that for $a, \beta \in (0, 1/2)$,

$$\underline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,odd}}(a, \beta) = \overline{\Gamma}^{(\infty)}_{\mathrm{Ball,\,odd}}(a, \beta) = \overline{\Gamma}^{(\infty)}(a, a). \tag{11.8}$$

The ball-noise stability problem can be interpreted as an isoperimetric problem in the r^{th} power of the Hamming graph [194]. The edge-isoperimetric inequality in Theorem 9.5 is a special case of this isoperimetric problem with $r = 1$.

In addition, similar questions can be posed when we replace the ball-noise with the sphere-noise. That is, we keep all things unchanged apart from the fact that $Z^n \sim \mathrm{Unif}(\mathbb{B}_r)$ is replaced by $Z^n \sim \mathrm{Unif}(\mathbb{S}_r)$, where $\mathbb{S}_r$ is the Hamming sphere centered at 0^n (cf. Section 8.2.3). For this case, the equalities for the odd case in (11.8) still holds. However, Yu [194] conjectured that the term $\overline{\Gamma}^{(\infty)}(a, a)$ in (11.7) should be replaced by $\frac{1}{2}\overline{\Gamma}^{(\infty)}(2a, 2a)$. For more details, please refer to [194].

Acknowledgements

We sincerely thank the anonymous reviewers for their careful reading and their many insightful comments and suggestions. We would also like to thank the Editor-in-Chief Professor Alexander Barg, and Mr. Mike Casey from Now Publishers for their advice in preparing the monograph. We are extremely grateful to our colleagues Zhaoqiang Liu, Anshoo Tandon, Junwen Yang, Qiaosheng Zhang, and especially Lin Zhou for their help in proofreading parts of the monograph.

Lei Yu is supported by the National Natural Science Foundation of China (NSFC) grant 62101286 and the Fundamental Research Funds for the Central Universities of China (Nankai University).

Vincent Tan is supported by a Singapore National Research Foundation (NRF) Fellowship (A-0005077-00-00) and Singapore Ministry of Education AcRF Tier 1 grants (A-0009042-00-00, A-8000189-00-00, and A-8000196-00-00). He would like to thank his wife Huili Guo and his four children Oliver Tan Ying Ren, Giselle Tan Ying Ci, Hazel Tan Ying Shan, and Ashleigh Tan Ying Xi for their unwavering support and understanding during the writing of this monograph.

References

[1] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Annals of Probability*, 1976, pp. 925–939.

[2] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," in *General Theory of Information Transfer and Combinatorics*, Springer, 2006, pp. 664–677.

[3] Y. Altuğ and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Transactions on Information Theory*, vol. 60, no. 8, 2014, pp. 4417–4426.

[4] V. Anantharam, "A variational characterization of Rényi divergences," *IEEE Transactions on Information Theory*, vol. 64, no. 11, 2018, pp. 6979–6989.

[5] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and a data processing inequality," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 3022–3026, Honolulu, Hawaii, USA, 2014.

[6] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, vol. 10. Société mathématique de France Paris, 2000.

[7] D. Bakry, "L'hypercontractivité et son utilisation en théorie des semigroupes," in *Lectures on Probability Theory*, Springer, 1994, pp. 1–114.

[8] D. Bakry, "Functional inequalities for Markov semigroups," in *Probability Measures on Groups*, Tata Institute of Fundamental Research, Mumbai, pp. 91–147, 2004.

[9] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 348. Springer Science & Business Media, 2013.

[10] D. Bakry and M. Ledoux, "Lévy–Gromov's isoperimetric inequality for an infinite dimensional diffusion generator," *Inventiones Mathematicae*, vol. 123, no. 2, 1996, pp. 259–281.

[11] L. P. Barnes and A. Özgür, "The Courtade–Kumar most informative Boolean function conjecture and a symmetrized Li–Médard conjecture are equivalent," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 2205–2209, Los Angeles, California, USA, 2020.

[12] F. Barthe, "On a reverse form of the Brascamp–Lieb inequality," *Inventiones Mathematicae*, vol. 134, no. 2, 1998, pp. 335–361.

[13] L. B. Beasley and T. J. Laffey, "Real rank versus nonnegative rank," *Linear Algebra and its Applications*, vol. 431, no. 12, 2009, pp. 2330–2335.

[14] S. Beigi and A. Gohari, "Quantum achievability proof via collision relative entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 12, 2014, pp. 7980–7986.

[15] S. Beigi and A. Gohari, "Φ-entropic measures of correlation," *IEEE Transactions on Information Theory*, vol. 64, no. 4, 2018, pp. 2193–2211.

[16] S. Beigi and C. Nair, "Equivalent characterization of reverse Brascamp–Lieb-type inequalities using information measures," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1038–1042, Barcelona, Spain, 2016.

[17] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Transactions on Information Theory*, vol. 60, no. 3, 2014, pp. 2926–2959.

[18] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, 2002, pp. 2637–2655.

[19] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, 1941, pp. 122–136.

[20] C. Bleuler, A. Lapidoth, and C. Pfister, "Conditional Rényi divergences and horse betting," *Entropy*, vol. 22, no. 3, 2020, p. 316.

[21] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[22] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, 2013, pp. 8077–8098.

[23] S. G. Bobkov, G. P. Chistyakov, and F. Götze, "Rényi divergence and the central limit theorem," *Annals of Probability*, vol. 47, no. 1, 2019, pp. 270–323.

[24] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Transactions on Information Theory*, vol. 57, no. 10, 2011, pp. 6351–6355.

[25] A. Bonami, "Ensembles $\Lambda(p)$ dans le dual de D^∞," in *Annales de l'institut Fourier*, vol. 18, pp. 193–204, 1968.

[26] A. Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(G)$," in *Annales de l'institut Fourier*, vol. 20, pp. 335–402, 1970.

[27] C. Borell, "Positivity improving operators and hypercontractivity," *Mathematische Zeitschrift*, vol. 180, no. 3, 1982, pp. 225–234.

[28] C. Borell, "Geometric bounds on the Ornstein–Uhlenbeck velocity process," *Probability Theory and Related Fields*, vol. 70, no. 1, 1985, pp. 1–13.

[29] 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 Mathematics*, vol. 20, no. 2, 1976, pp. 151–173.

[30] G. Braun, R. Jain, T. Lee, and S. Pokutta, "Information-theoretic approximations of the nonnegative rank," *Computational Complexity*, vol. 26, 2017, pp. 147–197.

[31] G. Braun and S. Pokutta, "Common information and unique disjointness," in *IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 688–697, 2013.

[32] C. Cai and S. Verdú, "Conditional Rényi divergence saddlepoint and the maximization of α-mutual information," *Entropy*, vol. 21, no. 961, 2019.

[33] E. A. Carlen and D. Cordero-Erausquin, "Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities," *Geometric and Functional Analysis*, vol. 19, no. 2, 2009, pp. 373–405.

[34] A. Chakrabarti and O. Regev, "An optimal lower bound on the communication complexity of gap-Hamming-distance," *SIAM Journal on Computing*, vol. 41, no. 5, 2012, pp. 1299–1317.

[35] M.-C. Chang, "A polynomial bound in Freiman's theorem," *Duke Mathematical Journal*, vol. 113, no. 3, 2002, pp. 399–419.

[36] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation.* Wiley, 2009.

[37] E. Çınlar, *Probability and Stochastics*, vol. 261. Springer, 2011.

[38] J. E. Cohen and U. G. Rothblum, "Nonnegative ranks, decompositions, and factorizations of nonnegative matrices," *Linear Algebra and its Applications*, vol. 190, no. 1, 1993, pp. 149–168.

[39] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd. McGraw-Hill Science/Engineering/Math, 2003.

[40] T. A. Courtade and G. R. Kumar, "Which Boolean functions maximize mutual information on noisy inputs?" *IEEE Transactions on Information Theory*, vol. 60, no. 8, 2014, pp. 4515–4525.

[41] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Transactions on Information Theory*, vol. 21, no. 3, 1975, pp. 226–228.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd. Wiley-Interscience, 2006.

[43] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, 1995, pp. 26–34.

[44] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, 1978, pp. 339–348.

[45] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[46] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Transactions on Information Theory*, vol. 46, no. 2, 2000, pp. 344–366.

[47] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Transactions on Information Theory*, vol. 57, no. 8, 2011, pp. 5509–5523.

[48] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, 2013, pp. 7071–7096.

[49] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd. Springer, 1998.

[50] Y. Dodis and Y. Yu, "Overcoming weak expectations," in *Theory of Cryptography*, Springer, pp. 1–22, Berlin, Germany, 2013.

[51] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.

[52] R. Eldan, "A two-sided estimate for the Gaussian noise stability deficit," *Inventiones Mathematicae*, vol. 201, no. 2, 2015, pp. 561–624.

[53] E. Erkip, "The efficiency of information in investment," Ph.D. dissertation, Ph.D. dissertation, Dept. Electr. Eng., Stanford Univ. Press, Stanford, CA, USA, 1996.

[54] C.-G. Esseen, "On the Liapunoff limit of error in the theory of probability," *Arkiv För Matematik, Astronomi och Fysik*, vol. A28, no. 1, 1942, pp. 1–19.

[55] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. R. R. Thomas, "On the nonnegative rank of distance matrices," *Mathematical Programming*, vol. 153, no. 1, 2015, pp. 133–177.

[56] H. Fawzi and P. Parrilo, "Lower bounds on nonnegative rank via nonnegative nuclear norms," *Mathematical Programming Series B*, vol. 153, no. 1, 2015, pp. 41–66.

[57] M. Fekete, "Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten," *Mathematische Zeitschrift*, vol. 17, no. 1, 1923, pp. 228–249.

[58] Y. Filmus, H. Hatami, S. Heilman, E. Mossel, R. O'Donnell, S. Sachdeva, A. Wan, and K. Wimmer, "Real analysis in computer science: A collection of open problems," *Preprint available at https://simons.berkeley.edu/sites/default/files/openprobsmerged.pdf*, 2014.

[59] F.-W. Fu, V. K. Wei, and R. W. Yeung, "On the minimum average distance of binary codes: Linear programming approach," *Discrete Applied Mathematics*, vol. 111, no. 3, 2001, pp. 263–281.

[60] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, 1973, pp. 149–162.

[61] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge: Cambridge University Press; 2014.

[62] M. Gastpar and E. Suha, "Relaxed Wyner's common information," in *IEEE Information Theory Workshop (ITW)*, pp. 1–5, Visby, Gotland, Sweden, 2019.

[63] H. Gebelein, "Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, 1941, pp. 364–379.

[64] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences," *IEEE Transactions on Information Theory*, vol. 56, no. 11, 2010, pp. 5377–5387.

[65] N. Gillis, *Nonnegative Matrix Factorization*. Society for Industrial & Applied Mathematics, 2020.

[66] R. Graczyk and A. Lapidoth, "Gray–Wyner and Slepian–Wolf guessing," in *IEEE International Symposium on Information Theory (ISIT)*, IEEE, pp. 2207–2211, Los Angeles, California, USA, 2020.

[67] R. Graczyk, A. Lapidoth, and M. Wigger, "Conditional and relevant common information," *Information and Inference: A Journal of the IMA*, vol. iaab021, 2022.

[68] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *The Bell Systems Technical Journal*, vol. 53, 1974, pp. 1681–1721.

[69] L. Gross, "Logarithmic Sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, 1975, pp. 1061–1083.

[70] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref, "Channel simulation via interactive communications," *IEEE Transactions on Information Theory*, vol. 63, no. 5, 2017, pp. 2659–2677.

[71] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, 2003.

[72] T. S. Han, "Weak variable-length source coding," *IEEE Transactions on Information Theory*, vol. 46, no. 4, 2006, pp. 1217–1226.

[73] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, 1993, pp. 752–772.

[74] L. H. Harper, "Optimal assignments of numbers to vertices," *Journal of the Society for Industrial and Applied Mathematics*, vol. 12, no. 1, 1964, pp. 131–135.

[75] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, 2010, pp. 438–449.

[76] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 52, no. 4, 2006, pp. 1562–1575.

[77] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Transactions on Information Theory*, vol. 54, no. 10, 2008, pp. 4619–4637.

[78] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, 2009, pp. 4947–4966.

[79] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, 2011, pp. 3989–4001.

[80] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, 2003, pp. 1753–1768.

[81] M. Hayashi and V. Y. F. Tan, "Equivocations, exponents and second-order coding rates under various Rényi information measures," *IEEE Transactions on Information Theory*, vol. 63, no. 2, 2017, pp. 975–1005.

[82] S. Heilman, "Euclidean partitions optimizing noise stability," *Electronic Journal of Probability*, vol. 19, no. 71, 2014, pp. 1–37.

[83] S. Heilman, "Low correlation noise stability of symmetric sets," *Journal of Theoretical Probability*, vol. 34, no. 4, 2021, pp. 2192–2240.

[84] H. O. Hirschfeld, "A connection between correlation and contingency," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 31, pp. 520–524, 1935.

[85] P. Hrubeš, "On the nonnegative rank of distance matrices," *Information Processing Letters*, vol. 112, 2012, pp. 457–461.

[86] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, 1952, pp. 1098–1101.

[87] M. Isaksson and E. Mossel, "Maximally stable Gaussian partitions with discrete applications," *Israel Journal of Mathematics*, vol. 189, no. 1, 2012, pp. 347–396.

[88] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional Rényi entropies," in *International Conference on Information Theoretic Security (ICITS)*, pp. 103–121, Singapore, 2013.

[89] R. Jain, Y. Shi, Z. Wei, and S. Zhang, "Efficient protocols for generating bipartite classical distributions and quantum states," *IEEE Transactions on Information Theory*, vol. 59, no. 8, 2013, pp. 5171–5178.

[90] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on Boolean functions," in *IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 68–80, 1988.

[91] G. Kalai and N. Linial, "On the distance distribution of codes," *IEEE Transactions on Information Theory*, vol. 41, no. 5, 1995, pp. 1467–1472.

[92] S. Kamath, "Reverse hypercontractivity using information measures," in *Allerton Conference on Communication, Control, and Computing*, pp. 627–633, Monticello, Illinois, USA, 2015.

[93] S. Kamath and V. Anantharam, "A new dual to the Gács-Körner common information defined via the Gray-Wyner system," in *Allerton Conference on Communication, Control, and Computing*, pp. 1340–1346, Monticello, IL, 2010.

[94] S. Kamath and V. Anantharam, "On non-interactive simulation of joint distributions," *IEEE Transactions on Information Theory*, vol. 62, no. 6, 2016, pp. 3419–3435.

[95] K. Kiener, "Uber produkte von quadratisch integrierbaren funktionen endlicher vielfalt," Ph.D. dissertation, Universität Innsbruck, 1969.

[96] G. Kindler, R. O'Donnell, and D. Witmer, "Remarks on the most informative function conjecture at fixed mean", 2015. eprint: arXiv:1506.03167.

[97] N. Kirshner and A. Samorodnitsky, "A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres," *IEEE Transactions on Information Theory*, vol. 67, no. 6, 2021, pp. 3509–3541.

[98] H. Koga and H. Yamamoto, "Asymptotic properties on codeword lengths of an optimal FV code for general sources," *IEEE Transactions on Information Theory*, vol. 51, no. 4, 2005, pp. 1546–1555.

[99] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *6th Prague Conference on Information Theory*, pp. 411–425, 1973.

[100] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 9, 2015, pp. 4316–4330.

[101] L. G. Kraft, "A device for quantizing, grouping, and coding amplitude modulated pulses," M.S. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1949.

[102] G. R. Kumar and T. A. Courtade, "Which Boolean functions are most informative?" In *IEEE International Symposium on Information Theory (ISIT)*, pp. 226–230, Istanbul, Turkey, 2013.

[103] G. R. Kumar, C.-T. Li, and A. El Gamal, "Exact common information," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 161–165, Honolulu, Hawaii, USA, 2014.

[104] E. N. Laguerre, *Théorie des Équations Numériques*. Gauthier-Villars, 1884.

[105] M. Ledoux, *Concentration of Measure and Logarithmic Sobolev Inequalities*, vol. 1709, ser. Séminaire de Probabilités XXXIII. Lecture Notes in Mathematics. Springer, Berlin, Heidelberg, 2006.

[106] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, 1999, pp. 788–791.

[107] C.-T. Li and A. El Gamal, "Distributed simulation of continuous random variables," *IEEE Transactions on Information Theory*, vol. 63, no. 10, 2017, pp. 6329–6343.

[108] C.-T. Li and A. El Gamal, "Extended Gray–Wyner system with complementary causal side information," *IEEE Transactions on Information Theory*, vol. 64, no. 8, 2017, pp. 5862–5878.

[109] C.-T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, 2018, pp. 6967–6978.

[110] J. Li and M. Médard, "Boolean functions: Noise stability, non-interactive correlation distillation, and mutual information," *IEEE Transactions on Information Theory*, vol. 67, no. 2, 2021, pp. 778–789.

[111] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends® in Communications and Information Theory*, vol. 5, no. 4-5, 2019, pp. 355–580, DOI: http://dx.doi.org/10.1561/0100000036.

[112] J. Liu, "Information theory from a functional viewpoint," Ph.D. dissertation, Ph.D. dissertation, Dept. Electr. Eng., Princeton, NJ: Princeton University, 2018.

[113] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Brascamp–Lieb inequality and its reverse: An information theoretic view," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1048–1052, Barcelona, Spain, 2016.

[114] W. Liu, G. Xu, and B. Chen, "The common information of N dependent random variables," in *Allerton Conference on Communication, Control, and Computing*, pp. 836–843, Monticello, Illinois, USA, 2010.

[115] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Transactions on Information Theory*, vol. 60, no. 5, 2014, pp. 2856–2867.

[116] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *IEEE Information Theory Workshop (ITW)*, pp. 501–505, Hobart, Tasmania, Australia, 2014.

[117] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Advances in Cryptology (EUROCRYPT)*, pp. 351–368, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[118] A. Moitra, "An almost optimal algorithm for computing nonnegative rank," *SIAM Journal of Computing*, vol. 45, no. 1, 2016, pp. 156–173.

[119] E. Mossel, "Mixing in product spaces," 2017, URL: http://math.mit.edu/~elmos/slides.pdf.

[120] E. Mossel, "Probabilistic view of voting, paradoxes, and manipulation," *Bull. Amer. Math. Soc.*, 2021.

[121] E. Mossel and J. Neeman, "Robust optimality of Gaussian noise stability," *Journal of the European Mathematical Society*, vol. 17, no. 2, 2015, pp. 433–482.

[122] E. Mossel and R. O'Donnell, "Coin flipping from a cosmic source: On error correction of truly random bits," *Random Structures & Algorithms*, vol. 26, no. 4, 2005, pp. 418–436.

[123] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, "Noise stability of functions with low influences: Invariance and optimality," *Annals of Mathematics*, vol. 171, no. 1, 2010, pp. 295–341.

[124] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality," *Israel Journal of Mathematics*, vol. 154, no. 1, 2006, pp. 299–336.

[125] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Geometric and Functional Analysis*, vol. 23, no. 3, 2013, pp. 1062–1097.

[126] C. Nair, "Equivalent formulations of hypercontractivity using information measures," in *International Zurich Seminar (IZS) Workshop*, Zürich, Switzerland, 2014.

[127] C. Nair and Y. N. Wang, "Evaluating hypercontractivity parameters using information measures," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 570–574, Barcelona, Spain, 2016.

[128] C. Nair and Y. N. Wang, "Reverse hypercontractivity region for the binary erasure channel," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 938–942, Aachen, Germany, 2017.

[129] J. O. Neeman, "Isoperimetry and noise sensitivity in Gaussian space," Ph.D. dissertation, University of California, Berkeley, 2013.

[130] R. O'Donnell, "Open problems in analysis of Boolean functions", 2012. eprint: arXiv:1204.6447.

[131] R. O'Donnell, *Analysis of Boolean Functions.* Cambridge University Press, 2014.

[132] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, 2018, p. 352.

[133] O. Ordentlich, Y. Polyanskiy, and O. Shayevitz, "A note on the probability of rectangles for correlated binary strings," *IEEE Transactions on Information Theory*, vol. 66, no. 11, 2020, pp. 7878–7886.

[134] O. Ordentlich, O. Shayevitz, and O. Weinstein, "An improved upper bound for the most informative Boolean function conjecture," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 500–504, Barcelona, Spain, 2016.

[135] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, 2001, pp. 903–917.

[136] G. Pichler, P. Piantanida, and G. Matz, "Dictator functions maximize mutual information," *The Annals of Applied Probability*, vol. 28, no. 5, 2018, pp. 3094–3101.

[137] Y. Polyanskiy, "Hypercontractivity for sparse functions on the discrete hypercube," *Manuscript*, 2019.

[138] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, 2010, pp. 2307–2359.

[139] Y. Polyanskiy and A. Samorodnitsky, "Improved log-Sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube," *Journal of Functional Analysis*, vol. 277, no. 11, 2019.

[140] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *Allerton Conference on Communication, Control, and Computing*, pp. 1334–1339, Monticello, Illinois, USA, 2010.

[141] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Theory.* New York, NY, USA: Springer-Verlag, 1998.

[142] M. Raginsky, "Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels," *IEEE Transactions on Information Theory*, vol. 62, no. 6, 2016, pp. 3355–3389.

[143] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications and coding," *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 1-2, 2013, pp. 1–246, DOI: http://dx.doi.org/10.1561/0100000064.

[144] A. Rényi, "On measures of dependence," *Acta Mathematica Hungarica*, vol. 10, no. 3-4, 1959, pp. 441–451.

[145] R. Rudnicki, M. Pichór, and M. Tyran-Kamińska, *Markov Semigroups and Their Applications*, vol. 597, ser. Dynamics of Dissipation. Lecture Notes in Physics. Springer, Berlin, Heidelberg, 2002.

[146] Y. Sakai, R. C. Yavas, and V. Y. F. Tan, "Third-order asymptotics of variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 67, no. 12, 2021, pp. 7708–7722.

[147] S. Salamatian, A. Cohen, and M. Médard, "Approximate Gács-Körner common information," in *IEEE International Symposium on Information Theory (ISIT)*, IEEE, pp. 2234–2239, Los Angeles, California, USA, 2020.

[148] A. Samorodnitsky, "A modified logarithmic Sobolev inequality for the Hamming cube and some applications", 2008. eprint: arXiv:0807.1679.

[149] A. Samorodnitsky, "On the entropy of a noisy function," *IEEE Transactions on Information Theory*, vol. 62, no. 10, 2016, pp. 5446–5464.

[150] I. Sanov, "On the probability of large deviations of random variables," *Mat. Sbornik*, no. 1, 1961, pp. 11–44.

[151] I. Sason, "On the Rényi divergence, joint range of relative entropies, and a channel coding theorem," *IEEE Transactions on Information Theory*, vol. 62, no. 1, 2016, pp. 23–34.

[152] M. Schreiber, "Fermeture en probabilité de certains sous-espaces d'un espace L^2," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, 1969, pp. 36–48.

[153] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell Systems Technical Journal*, vol. 27, 1948, pp. 379–423.

[154] O. Shayevitz, "On Rényi measures and hypothesis testing," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 894–898, St Petersburg, Russia, 2011.

[155] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitstheorie verw. Geb*, vol. 14, 1969, pp. 149–160.

[156] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, 1973, pp. 471–80.

[157] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Transactions on Information Theory*, vol. 42, no. 1, 1996, pp. 63–86.

[158] M. Sudan, H. Tyagi, and S. Watanabe, "Communication for generating correlation: A unifying survey," *IEEE Transactions on Information Theory*, vol. 66, no. 1, 2020, pp. 5–37.

[159] V. Y. F. Tan, "Moderate-deviations of lossy source coding for discrete and Gaussian sources," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 920–924, Cambridge, MA, 2012.

[160] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1-2, 2014, pp. 1–184, DOI: http://dx.doi.org/10.1561/0100000086.

[161] V. Y. F. Tan and M. Hayashi, "Analysis of remaining uncertainties and exponents under various conditional Rényi entropies," *IEEE Transactions on Information Theory*, vol. 64, no. 5, 2018, pp. 3734–3755.

[162] V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the β-divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, 2013, pp. 1592–1605.

[163] H. Thorisson, *Coupling, Stationarity, and Regeneration*, vol. 14. Springer New York, 2000.

[164] C. Tsallis, "What are the numbers that experiments provide," *Quimica Nova*, vol. 17, no. 6, 1994, pp. 468–471.

[165] H. Tyagi, "Common information and secret key capacity," *IEEE Transactions on Information Theory*, vol. 59, no. 9, 2013, pp. 5627–5640.

[166] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, 2014, pp. 3797–3820.

[167] A. Vandaele, N. Gillis, F. Glineur, and D. Tuyttens, "Heuristics for exact nonnegative matrix factorization," *Journal of Global Optimization*, vol. 65, 2016, pp. 369–400.

[168] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM Journal of Optimization*, vol. 20, no. 3, 2009, pp. 1364–1377.

[169] B. N. Vellambi and J. Kliewer, "Sufficient conditions for the equality of exact and Wyner common information," in *Allerton Conference on Communication, Control, and Computing*, pp. 370–377, Monticello, Illinois, USA, 2016.

[170] B. N. Vellambi and J. Kliewer, "New results on the equality of exact and Wyner common information rates," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 151–155, Vail, Colorado, USA, 2018.

[171] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Transactions on Information Theory*, vol. 41, no. 5, 1995, pp. 1322–1332.

[172] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, 1994, pp. 1147–1157.

[173] K. B. Viswanatha, E. Akyol, and K. Rose, "The lossy common information of correlated sources," *IEEE Transactions on Information Theory*, vol. 60, no. 6, 2014, pp. 3238–3253.

[174] C.-Y. Wang, S. H. Lim, and M. Gastpar, "Information-theoretic caching: Sequential coding for computing," *IEEE Transactions on Information Theory*, vol. 62, no. 11, 2016, pp. 6393–6406.

[175] A. Winter, "Compression of sources of probability distributions and density operators", 2002. eprint: arXiv:0208131.

[176] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 5, 1975, pp. 493–501.

[177] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM Journal on Applied Mathematics*, vol. 31, no. 2, 1976, pp. 313–333.

[178] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, 1975, pp. 100–113.

[179] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, 1957, pp. 591–606.

[180] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd. Springer-Verlag, New York, 1978.

[181] L. Wu, "Large deviations, moderate deviations and LIL for empirical processes," *Annals of Probability*, 1994, pp. 17–27.

[182] A. D. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, 1975, pp. 163–179.

[183] A. D. Wyner, "The wire-tap channel," *The Bell Systems Technical Journal*, vol. 54, 1975, pp. 1355–1387.

[184] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Transactions on Information Theory*, vol. 19, no. 6, 1973, pp. 769–772.

[185] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, 1976, pp. 1–10.

[186] G. Xu, W. Liu, and B. Chen, "A lossy source coding interpretation of Wyner's common information," *IEEE Transactions on Information Theory*, vol. 62, no. 2, 2016, pp. 754–768.

[187] H. Yang and R. D. Wesel, "On the most informative Boolean functions of the very noisy channel," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1202–1206, Paris, France, 2019.

[188] K. Yang, "On the (im)possibility of non-interactive correlation distillation," *Theoretical Computer Science*, vol. 382, no. 2, 2007, pp. 157–166.

[189] M. Yannakakis, “Expressing combinatorial optimization problems by linear programs,” *Journal of Computer and System Sciences*, vol. 43, no. 4, 1991, pp. 441–466.

[190] M. H. Yassaee, A. Gohari, and M. R. Aref, “Channel simulation via interactive communications,” *IEEE Transactions on Information Theory*, vol. 61, no. 6, 2015, pp. 2964–2982.

[191] L. Yu, “*“On the Φ-stability and related conjectures”*,” 2021. eprint: arXiv:2104.08740.

[192] L. Yu, “Strong Brascamp–Lieb inequalities”, 2021. eprint: arXiv:2102.06935.

[193] L. Yu, “The convexity and concavity of envelopes of the minimum-eelative-entropy region for the DSBS”, 2021. eprint: arXiv:2106.03654.

[194] L. Yu, “Edge-isoperimetric inequalities and ball-noise stability: Linear programming and probabilistic approaches,” *Journal of Combinatorial Theory, Series A*, vol. 188, no. 105583, May 2022, pp. 1–33.

[195] L. Yu, V. Anantharam, and J. Chen, “Graphs of joint types, non-interactive simulation, and stronger hypercontractivity”, 2021. eprint: arXiv:2102.00668.

[196] L. Yu, H. Li, and C. W. Chen, “Generalized common informations: Measuring commonness by the conditional maximal correlation”, 2016. eprint: arXiv:1610.09289.

[197] L. Yu and V. Y. F. Tan, “Wyner’s common information under Rényi divergence measures,” *IEEE Transactions on Information Theory*, vol. 64, no. 5, 2018, pp. 3616–3623.

[198] L. Yu and V. Y. F. Tan, “An improved linear programming bound on the average distance of a binary code”, 2019. eprint: arXiv:1910.09416.

[199] L. Yu and V. Y. F. Tan, “Asymptotic coupling and its applications in information theory,” *IEEE Transactions on Information Theory*, vol. 65, no. 3, 2019, pp. 1321–1344.

[200] L. Yu and V. Y. F. Tan, “Rényi resolvability and its applications to the wiretap channel,” *IEEE Transactions on Information Theory*, vol. 65, no. 3, 2019, pp. 1862–1897.

[201] L. Yu and V. Y. F. Tan, “Simulation of random variables under Rényi divergence measures of all orders,” *IEEE Transactions on Information Theory*, vol. 65, no. 6, 2019, pp. 3349–3383.

[202] L. Yu and V. Y. F. Tan, “Corrections to “Wyner’s common information under Rényi divergence measures”,” *IEEE Transactions on Information Theory*, vol. 66, no. 4, 2020, pp. 2599–2608.

[203] L. Yu and V. Y. F. Tan, “Exact channel synthesis,” *IEEE Transactions on Information Theory*, vol. 66, no. 5, 2020, pp. 2299–2818.

[204] L. Yu and V. Y. F. Tan, “On exact and ∞-Rényi common information,” *IEEE Transactions on Information Theory*, vol. 66, no. 6, 2020, pp. 3366–3406.

[205] L. Yu and V. Y. F. Tan, “On non-interactive simulation of binary random variables,” *IEEE Transactions on Information Theory*, vol. 67, no. 4, 2021, pp. 2528–2538.