ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUL 08 2013
LIBRARIES

Locally Decodable Source Coding

by

Ali Makhdoumi

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

© Massachusetts Institute of Technology 2013. All rights reserved.

Author .....
Department of Electrical Engineering and Computer Science
May 20, 2013

Certified by .....
Prof. Muriel Médard
EECS Professor
Thesis Supervisor

Certified by .....
Prof. Yury Polyanskiy
EECS Assistant Professor
Thesis Supervisor

Accepted by .....
Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Locally Decodable Source Coding

by

Ali Makhdoumi

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Electrical Engineering

Abstract

Source coding is accomplished via the mapping of consecutive source symbols (blocks) into code blocks of fixed or variable length. The fundamental limits in source coding introduces a tradeoff between the rate of compression and the fidelity of the recovery. However, in practical communication systems many issues such as computational complexity, memory capacity, and memory access requirements must be considered. In conventional source coding, in order to retrieve one coordinate of the source sequence, accessing all the encoded coordinates are required. In other words, querying all of the memory cells is necessary.
We study a class of codes for which the decoder is local. We introduce locally decodable source coding (LDSC), in which the decoder need not to read the entire encoded coordinates and only a few queries suffice to retrieve a given source coordinate. Both cases of having a constant number of queries and also a scaling number of queries with the source block length are studied. Also, both lossless and lossy source coding are considered. We show that with constant number of queries, the rate of (almost) lossless source coding is one, meaning that no compression is possible. We also show that with logarithmic number of queries in block length, one can achieve Shannon entropy rate. Moreover, we provide achievability bound on the rate of lossy source coding with both constant and scaling number of queries.

Thesis Supervisor: Prof. Muriel Médard
Title: EECS Professor

Thesis Supervisor: Prof. Yury Polyanskiy
Title: EECS Assistant Professor

Acknowledgments

I would like to thank my advisors Prof. Muriel Médard and Prof. Yury Polyanskiy. I value our discussions on different research challenges as well as their insightful guidance. They have been extremely supportive mentors, both professionally and personally. I also want to thank Mr. Shao-Lun Huang for his valuable comments on this thesis. I am also grateful to Prof. Pablo Parrilo, my academic mentor, for his valuable advices during my master study.

I could not have come this far without the love of my family. I am always grateful to my parents, my sisters, and my wonderful friends for theirs endless love and support.

Contents

1 Introduction **11**

1.1 Motivation 12

2 Background and Literature Review **15**

2.1 Source Coding 15

2.1.1 Almost Lossless Source Coding 15

2.1.2 Lossy Source Coding 17

2.2 Locally Encodable Source Coding 19

2.2.1 Lossless LESC 19

2.2.2 Lossy LESC 22

2.3 Succinct Date Structure 23

3 Locally Decodable Source Coding **25**

3.1 Lossless Locally Decodable Source Coding 26

3.1.1 LDSC 26

3.1.2 Average LDSC 28

3.1.3 Linear Encoder 30

3.1.4 Linear Decoder 33

3.1.5 General Encoder-Decoder 34

3.1.6 On the Bounded Degree Bipartite Graph 36

3.1.7 Scaling Number of Queries 38

3.2 Lossy Locally Decodable Source Coding 40

3.2.1 Scaling number of queries 41

3.2.2 Fixed Number of queries 44
3.3 LDLSC for Excess Distortion 45
3.4 Fixed to Variable Length Local Encoding-Decoding 47
3.5 Information Preserving Transformations 54
3.6 Storage on Memories with Access Cost 56

4 Conclusion and future work **61**

List of Figures

1-1 Typical structure of a communication system 12

2-1 Locally Encodable Source Coding 19

2-2 Locally encodable Source Coding: Definition of Locality 20

3-1 Locally Decodable Source Coding 26

3-2 Comparison of approximations of upper bounds, $p = .11$, $d = 1/15$ 43

3-3 Comparison of approximations of upper bounds, $p = .11$, $d = 1/20$ 44

3-4 Local Encoder-Decoder 48

Chapter 1

Introduction

The block structure of a typical communication system is illustrated in 1-1. The source generates a sequence X^n, the source encoder maps the sequence, X^n, into the bitstream, Y^k. The bitstream is transmitted over a possibly erroneous channel and the received bitstream $\hat{Y}^k$ is processed by the source decoder in order to produce the decoded source sequence $\hat{X}^n$.

The error probability of the channel is controlled by the channel encoder, which adds redundancy to the bits at the source encoder output, Y^k. Typically, there is a modulator and a demodulator. The modulator maps the channel encoder output to an analog signal, which is suitable for transmission over a physical channel. The demodulator interprets the received, often analog, signal as a digital signal, which is fed into the channel decoder. The channel decoder processes the digital signal and produces the received bitstream $\hat{Y}^k$, which may be identical to Y^k even in the presence of channel noise. According to separation results [23, 22], source coding and channel coding can be constructed separately without loss of throughput in the overall system. The focus of this work is on the source encoder and decoder parts. This part of communication system is called source coding. The main goal of source coding is to compress data source in a way to be recoverable with a high fidelity.

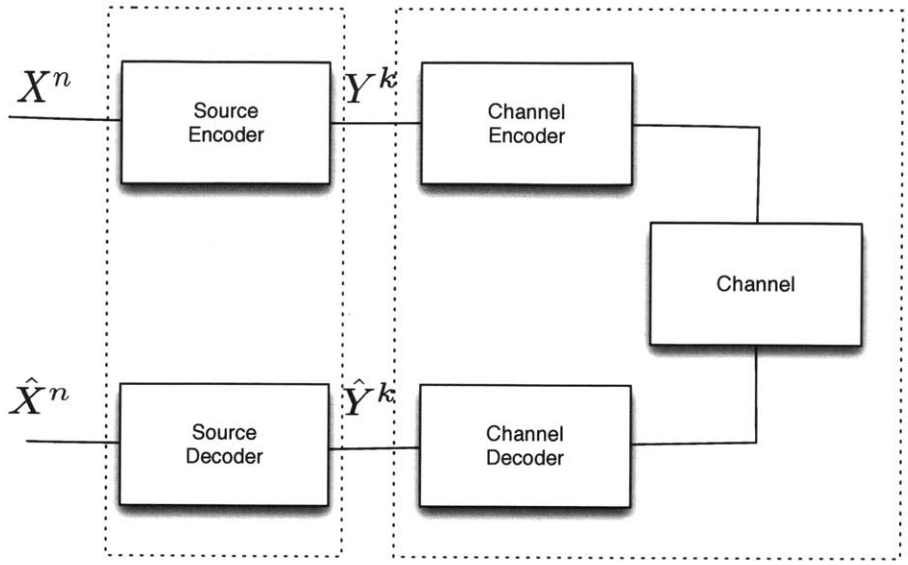

Figure 1-1: Typical structure of a communication system

1.1 Motivation

The basic communication problem may be expressed as transmitting source data with a high fidelity without exceeding an available bit rate, or it may be expressed as transmitting the source data using the lowest bit rate possible while maintaining a specified reproduction fidelity [23]. In either case, a fundamental trade-off is made between bit rate and distortion/error level. Source coding is primarily characterized by rate and distortion /error of the code. However, in practical communication systems, many issues such as *computational complexity*, *memory limitation*, *memory access requirements* must be considered. For instance, in a typical system, a small change in one coordinate of the input sequence leads to a large change in the encoded output. Moreover, in order to retrieve one symbol of the source sequence, accessing all the encoded coordinates are required. The latter issue is the main topic of this work.

One way to confront these issues is to place constraints on the encoder/decoder. In particular, in order to address the issue of memory access requirement, we study a class of codes for which the decoder is local (in a sense that will be defined later). A long line of research has addressed a similar problem from a data structure perspective. For example, Bloom filters [2] are a popular data structure for storing a set in a compressed form while

allowing membership queries to be answered in constant time. The rank/select problem [18, 8] and dictionary problem in the field of succinct data structures are also examples of problems involving both compression and the ability to efficiently recover a single element of the input. In particular, [20] gives a succinct data structure for arithmetic coding that supports efficient recovery of source. Reference [3] studies both issues at the same time and introduces a data structure with is efficient in both updating and querying. In all of these works the efficiency is interpreted in terms of the decoding time whereas in this work it is interpreted in terms of memory access requirement. In this work, we formulate this problem from an information theoretic view and study the fundamental trade-offs between locality and the rate of source coding.

A topic closely related to source coding with local decoding is the problem of source coding with local encoding. This problem has been studied in many works in both data structure and information theoretic literatures. This line of research addresses the following challenge: In order to be able to update an individual source symbol efficiently, we must study compression schemes that have some continuity property, meaning that a change in a single coordinate of the input sequence leads to a small change in the encoded sequence. Varshney et al. [25] analyzed continuous source codes from an information theoretic point of view . Also, Mossel and Montanari [16] have constructed source codes based on nonlinear sparse graph codes. Sparse linear codes has been studied by Mackay [13], in which a class of local linear encoders are introduced. Also, Mazumdar et. all. [15] has studied update efficient codes which studies channel coding problem with local encoders.

Causal Source Coding is another close topic to source coding with local decoder, which studies the source coding problem with causal encoder/decoder. In causal source coding the constraint on the decoder is not being local, but, being causal [17, 9].

Locally decodable codes (LDC) ([26]) is another close topic to source coding with local decoder. An LDC encodes n-bit source sequence to k-bit codewords in such a way that one can recover any bit x_i from a corrupted codeword by querying only a few bits of that codeword. Therefore, LDC introduces redundancy to combat the corruption of the code word.

Moreover, *Locally repairable codes* ([19]) study the case where the encoded coordinates,

y_is, face erasure with some erasure probability. The goal is to produce another y_i to replace a erased y_j, by accessing a few number of y_is. Reference [19] introduces a trade-off between locality, code distance, and the rate of code.

Chapter 2

Background and Literature Review

In this chapter we present some fundamental concepts of source coding. We introduce the class of *Locally Encodable Source Coding (LESC)* which studies the problem of source coding with a local encoder. Also, we overview the results of *Succinct Date Structure*. In the next chapter we shall revisit these results from an information theoretic point of view and compare them to our results.

2.1 Source Coding

The primary task of source coding is to represent a source with the minimum number of (binary) symbols without exceeding an acceptable level of distortion, which is determined by the application. Two types of source coding techniques are typically named almost lossless source coding and lossy source coding.

2.1.1 Almost Lossless Source Coding

Almost Lossless Source Coding refers to a type of source coding that allows the exact reconstruction of the original source from the compressed data for almost all the source outputs. Almost lossless source coding are also called asymptotically lossless source coding or some times by the abuse of name we call them lossless coding. Although that lossless compression such as Lempel-Ziv [28] exists and is generally markedly different in

construction. Lossless source coding can provide a reduction in bit rate compared to the original source, when the original data source contains dependencies or statistical properties such as redundancy, sparsity, and correlation that can be exploited for data compression . A well-known use for this type of compression for picture and video sources is JPEG-LS. Next, a fundamental bound for the minimum average codeword length per source symbol that can be achieved with lossless coding is introduced. Let $X \in \mathcal{X}$ be a random variable with probability measure $\mathbb{P}_X$, where $\mathcal{X}$ is a finite alphabet set. Also let X^n denotes n i.i.d copies of X.

Definition 1. *An (n, k, ϵ)-SC is a pair consist of encoder $f : X^n \to \{0,1\}^k$ and decoder $g : \{0,1\}^k \to X^n$ such that*

$$\mathbb{P}[g(f(X^n)) \neq X^n] \leq \epsilon.$$

Also let

$$k^*_{SC}(n, \epsilon) \triangleq \min\{k \ : \ \exists \ (n, k, \epsilon) - SC\},$$

$$R_{SC}(n, \epsilon) \triangleq \frac{k^*_{SC}(n, \epsilon)}{n},$$

and

$$R_{SC}(\epsilon) \triangleq \limsup_{n \to \infty} R_{SC}(n, \epsilon).$$

The rate is

$$R_{SC} \triangleq \lim_{\epsilon \to 0} R_{SC}(\epsilon),$$

where SC stands for source coding.

For any X with probability measure $\mathbb{P}_X$, we have

$$R_{SC} = H(X), \tag{2.1}$$

where $H(X)$ denotes the entropy of a random variable ([4], chapter 5). The following theorem from [11, 24] characterizes the finite block length results on the rate.

Theorem 1. *([11, 24]) For any X with probability measure $\mathbb{P}_X$, we have*

$$R_{SC}(n,\epsilon) = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon) + \Theta(\frac{\log n}{n}), \tag{2.2}$$

where $V(V) = Var(\log P_X(X))$ and Q^{-1} is the functional inverse of the Q-function, where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{y^2}{2}}dy$.

2.1.2 Lossy Source Coding

Lossy source coding refers to a type of source coding where a source is represented by a loss of information. In this case, only an approximation of the original source can be reconstructed from the compressed data. Lossy coding is the primary coding type for the compression of speech, audio, picture, and video signals, where an exact reconstruction of the source data is not required. A well-known application of lossy coding techniques is JPEG. A measure of the quality of the approximation, is referred to as the distortion.

The minimum number of bits per source symbol that are required for representing a given source without exceeding a given distortion level is called rate distortion.

To measure the quality of an approximation, distortion measures are defined to express the differences between a reconstructed source and the corresponding original data source as a non-negative real value. A smaller distortion corresponds to a higher approximation quality. A distortion of zero specifies that the reproduced symbols are identical to the corresponding original symbols. In this work, we restrict our considerations to the important class of additive distortion measures. The distortion between a single reconstructed symbol $\hat{x} \in \hat{\mathcal{X}}$ and the corresponding original symbol $x \in \mathcal{X}$ is dened as a function $d(x,\hat{x}) \geq 0$, with equality if and only if $x = \hat{x}$. Given such a distortion measure $d(x,\hat{x})$, the distortion between a sequence $\hat{x}^n$ and x^n is defined as

$$d(x^n,\hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i).$$

For binary sources where $\mathcal{X} = \hat{\mathcal{X}} = \{0,1\}$, the most commonly used additive distortion measure is the indicator distortion defined as $\mathbf{1}\{x \neq \hat{x}\}$. Next, we formally define the rate

of an average distortion code for a given source.

Definition 2. *A $(n, k, D)-$LSC is a pair consist of an encoder $f : X^n \to \{0,1\}^n$ and a decoder $g : \{0,1\}^k \to X^n$ such that*

$$\mathbb{E}[d(X^n, \hat{X}^n)] \leq D.$$

Let

$$k^*_{lsc}(n, D) \triangleq \min\{k \ : \ \exists\, (n, k, D) - LSC\},$$

$$R_{lsc}(n, D) \triangleq \frac{k^*_{lsc}(n, D)}{n},$$

and the rate is

$$R_{lsc}(D) \triangleq \limsup_{n\to\infty} R_{lsc}(n, D),$$

where lsc stands for lossy source coding.

For any X with probability measure $\mathbb{P}_X$, the characterization of the rate distortion is given by ([4], chapter 10)

$$R_{lsc}(D) = \min_{P_{\hat{X}|X} \ : \ \mathbb{E}[d(X,\hat{X})] \leq D} I(X; \hat{X}). \tag{2.3}$$

The following theorem characterizes the finite block length results on the rate of lossy source coding.

Theorem 2. *([27]) For any X with probability measure $\mathbb{P}_X$, we have*

$$R_{lsc}(n, d) \leq R_{lsc}(D) + \frac{\log n}{n} + o(\frac{\log n}{n}), \tag{2.4}$$

where $R_{lsc}(D)$ is given in (2.3).

We shall use the finite length results in the next chapters when we study locally decodable lossy source coding.

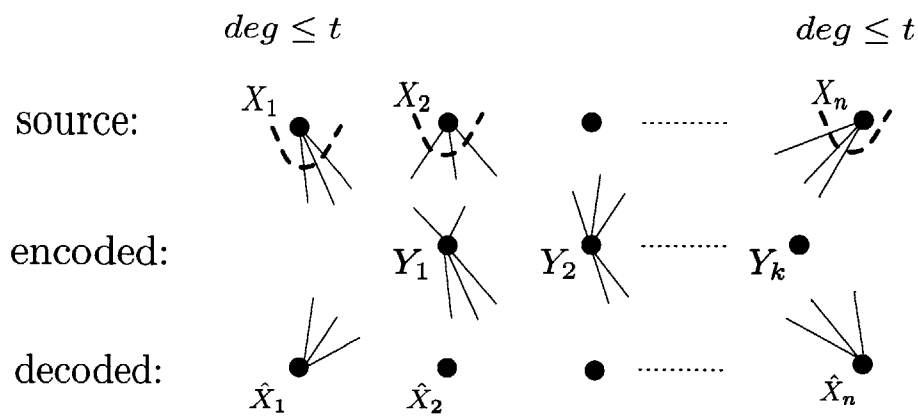

Figure 2-1: Locally Encodable Source Coding

2.2 Locally Encodable Source Coding

We term *Locally Encodable Source Coding (LESC)* source coding where the encoder is local (The name that we use is not found the literature). In particular, [13, 16] study the case where, for any i, X_i influences only a constant number (t) of Y_is. This class of source codes is depicted in Figure 2-1. In this figure, we connect Y_i to X_j, if X_j contributes to evaluate Y_i (we shall define this formally later). As shown in the figure, the source nodes have bounded degree determined by the locality. In this section, we formally define LESC and give the results in the literature for both lossless and lossy settings.

2.2.1 Lossless LESC

A lossless coding is defined as a pair of encoder f and decoder g, where, $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$. The encoder is called local if each coordinate of the input affects a bounded number of coordinates of output. Formally, Let f_a, for $a \in \{1, ..., k\}$, be the $a-$th component of the encoding function. Assume f_a depends on X^n only through the vector $X^{\mathcal{N}_a^Y} = \{X_j \ : \ j \in \mathcal{N}_a^Y\}$ for some $\mathcal{N}_a^Y \subset \{1, ..., n\}$. Also for $i \in \{1, ..., n\}$, let $\mathcal{N}_i^X$ be the set of output coordinates that depend on i. Thus,

$$\mathcal{N}_i^X = \{a \in \{1, ..., k\} \ : \ i \in \mathcal{N}_a^Y\}.$$

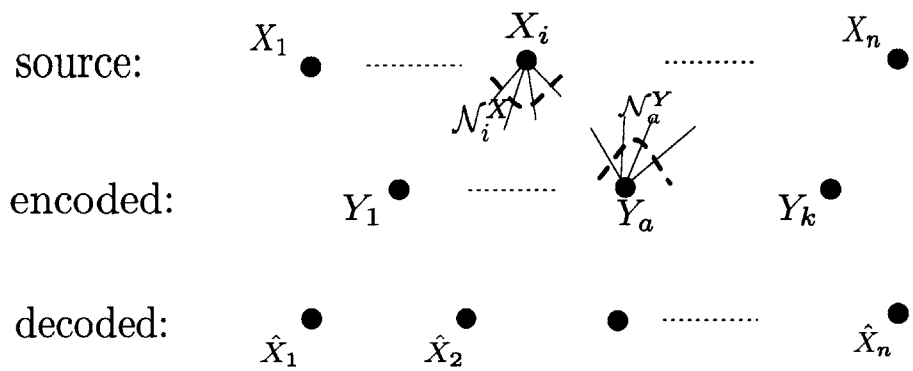

Figure 2-2: Locally encodable Source Coding: Definition of Locality

For any given t, an encoder is called $t-$local if $|\mathcal{N}_i^X| \leq t$ for any $i \in \{1, ..., n\}$. Figure 2-2 illustrates $\mathcal{N}_a^Y$ and $\mathcal{N}_i^X$.

Definition 3. *An $(n, k, t, \epsilon)-$LESC is a pair consisting a $t-$local encoder $f : X^n \mapsto \{0,1\}^k$ and a decoder $g : \{0,1\}^k \mapsto X^n$, such that $\mathbb{P}[g(f(X^n)) \neq X^n] \leq \epsilon$. Also, define*

$$k_{le}^*(n, t, \epsilon) \triangleq \min\{k \ : \ (n, k, t, \epsilon) - LESC\},$$

$$R_{le}(t, \epsilon) \triangleq \limsup_{n \to \infty} \frac{k_{le}^*(n, t, \epsilon)}{n},$$

and

$$R_{le}(t) \triangleq \lim_{\epsilon \to 0} R_{le}(t, \epsilon),$$

where the subscript le stands for local encoder.

Next, we present relevant results in the literature about source coding with a local encoder.

Assume a Bernoulli(p) i.i.d. source. The following result is derived in [13].

Theorem 3 (Mackay, *very good codes*, [13]). *Assume X is a Bern(p) i.i.d. source with entropy $h(p)$. For any given rate $R > h(p)$, there exists an integer $t(h(p), R) \geq 3$ such that for any desired block error $\epsilon > 0$, there exists an integer n_0 such that for any $n > n_0$ there exists a $(n, nR, t\,(h(p), R)\,, \epsilon)$ $-$LESC. Moreover, this encoder is a linear encoder.*

Note 1. *A Linear encoder for a Bernoulli source is a binary $n \times k$ matrix $\mathbf{G}$ such that $Y = X\mathbf{G}$ mod 2. This encoder is $t-$local if the weight of each row is at most t.*

Given a Bern(p) i.i.d. source and any rate $R > h(p)$, we have $R_{le}(t(h(p), R), \epsilon) \leq R$. Therefore, $R_{le}(t(h(p), R)) \leq R$. As a result, for any $R > h(p)$, there exists $t = t(h(p), R)$ such that $R_{le}(t) = R$. Thus, we obtain a rate arbitrarily close to entropy using LESC. But, in order to approach the entropy rate the locality must increase.

Source coding with local encoding is also studied in [16], considering a non-linear encoding function. In order to state the results we need the following definition. Let X be a source taking values from a finite set $\mathcal{X}$. For any integer $k \geq 1$, let $\mathcal{D}_k(\mathcal{X})$ denote the set of probability measures $\mathbb{P}_X$ over $\mathcal{X}$, such that there exists a function $f : \mathcal{X}^k \mapsto \mathcal{X}$ for which the following holds. If $X_1, ..., X_k$ are i.i.d with measure $\mathbb{P}_X$, then $f(X_1^n)$ is uniform in $\mathcal{X}$. Clearly, $\mathcal{D}_k(\mathcal{X})$ is finite and increasing in k, and $\mathcal{D}(\mathcal{X}) = \cup_k \mathcal{D}_k(\mathcal{X})$ is dense in the $|\mathcal{X} - 1|-$dimensional simplex of probability measures over $\mathcal{X}$.

The main result of [16] is as following.

Theorem 4 ([16]). *Let X be a an i.i.d source over $\mathcal{X}$ with probability measure $\mathbb{P}_X$. If $\mathbb{P}_X \in \mathcal{D}_k(\mathcal{X})$, then there exists $t^*(\mathbb{P}_X)$ such that for any $t \geq t^*(\mathbb{P}_X)$ we have*

$$H(\mathbb{P}_X) = \inf\{R \ : \ \exists \textit{ LESC with rate } R \textit{ and } t - \textit{local encoder}\}.$$

This theorem shows that, if the probability measure of the source comes from $\mathcal{D}(\mathcal{X})$ (a dense set on the space of probability distributions), then LESC achieves the fundamental limit, $H(X)$.

Note 2. *Unlike Theorem 3, the encoder proposed in Theorem 4 is not linear. Moreover, in Theorem 3, in order to approach entropy, we need to increase the locality t. In other words, the locality, t, is a function of both rate and probability measure, whereas in Theorem 4 the locality is only a function of probability measure.*

The number of queries, t can also be a growing function of the block length, n. We shall consider this case later.

2.2.2 Lossy LESC

Let $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto \hat{X}^n$ be encoder and decoder, respectively. The encoder is called local if each coordinate of the input affects a bounded number of coordinates of output. This definition is the same as the previous one. Note that $X \in \mathcal{X}$ and $\hat{X} \in \hat{\mathcal{X}}$. Assume a probability measure on $\mathcal{X}$, $\mathbb{P}_X$. A separable distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathbb{R}^+$ is given.

An $(n,k,t,D)-$LELSC is a pair consisting of a $t-$local encoder $f : X^n \mapsto \{0,1\}^k$ and a decoder $g : \{0,1\}^k \mapsto \hat{X}^n$, such that $\mathbb{E}[d(X^n, \hat{X}^n)] \leq D$. Also, define

$$k^*_{le}(n,t,D) \triangleq \min\{k \ : \ (n,k,t,D) - \text{LESC}\},$$

and

$$R_{le}(t,D) \triangleq \limsup_{n\to\infty} \frac{k^*_{le}(n,t,D)}{n}.$$

Next, we state the relevant results in the literature about lossy source coding with local encoder.

Dimakis et. al. [7] discuss the case where the source has an i.i.d. Bern($\frac{1}{2}$) distribution. The following theorem from [7] states that, if we choose the linear encoder generating matrix randomly, then with high probability the achievable rate by linear local encoder is bounded away from Shannon rate distortion. A Low Density Generating Matrices (LDGM) is used to guarantee locality. An LDGM is a $G_{n\times k}$ matrix that maps sequences of length n to sequences of length k. It has locality (low density) of t if each row of it has at most t non-zeros.

Theorem 5 ([7]). *Let X be a i.i.d source with $Bern(\frac{1}{2})$ distribution. Consider linear encoders that are chosen randomly from the set of all LDGMs with locality t . With high probability (wrt the ensemble) the achieved rate-distortion pair (R, D) satisfies:*

$$R_{le}(t,D) \geq (1-h(D))\frac{1}{1-\exp\left(-\frac{(1-D)t}{R_{le}(t,D)}\right)}. \tag{2.5}$$

Note that Theorem 5 gives a bound on the average rate that a randomly chosen code

can achieve and does not apply to individual codes. The authors of [12] generalized this result by using a counting argument to individual linear local encoders. They showed that, for any linear local encoder with locality t, the performance is strictly bounded away from the Shannon rate-distortion function. However, the rate approaches rate-distortion as t increases. Formally,

For any $R > 1-h(D)$, there exists $t(R)$, such that the rate R can be achieved with locality t.

Note that here, in order to approach the Shannon rate distortion function, we should increase the locality, t.

2.3 Succinct Date Structure

The problem of locally decodable (efficiently recoverable) compressed data structure (also called succinct data structure) is studied in many works from a database point of view. In succinct data structures we are concerned with the design of space efficient and dynamic data structures for storing a data source. We need the source data to be represented as compactly as possible to minimize storage, which is often highly constrained in these scenarios. On the other hand, we wish to retrieve any coordinate of data source efficiently . Therefore, we are interested in data structures for storing sequences of length n produced by a source in a compressed manner and being able to read the source coordinates efficiently. Reference [20, 3] study this problem and the authors develop space bounds for data compression, i.e. storing a sequence of integers in a compressed binary format. They additionally seek to design compressed data indices to decode any small portion of the data or search for any pattern as a substring of the data, without decompressing the binary stored sequence entirely. Patrascu, [20] considered the problem of mapping a sequence, X^n, into a sequence, Y^k, and recovering $\hat{X}^n$ from Y^k, where any $\hat{x}_i$ only depends on $O(\log n)$ of Y_is. This approach is very close to our locally decodable source coding scheme. We shall state the following result in the literature, in order to compare it to the results of LDSC in next chapter.

Theorem 6. *([20]) Consider a sequence of n elements from an alphabet $\mathcal{X}$, and let f_x be the number of occurrences of letter x in the sequence. On a RAM with cells of $(\log n)$ bits, we can represent the sequence by*

$$O(|\mathcal{X}| \log n) + \sum_{x \in \mathcal{X}} f_x \log(\frac{n}{f_x}) + \frac{n}{(\frac{\log n}{t})^t} + O(n^{3/4} \log n^{3/4})$$

bits of memory, supporting single-element access in $O(t)$. Each RAM is a sequence of bits with length $O(\log n)$. Single-element access in $O(t)$, means that each coordinate of the input sequence can be recovered by reading from $O(t)$ RAMs.

We shall elaborate on this result and reformulate it in a information theoretic form in the next chapter, where we shall develop the machinery to provide information theoretic explanation of succinct data structure, and in particular of Theorem 6.

Chapter 3

Locally Decodable Source Coding

The problem of *locally decodable source coding (LDSC)* is studied in this chapter where a source sequence $x_1, x_2, ...$, taking values from the source alphabet $\mathcal{X}$, is mapped into a sequence $y_1, y_2, ...$ of symbols taking values in the compression alphabet $\mathcal{Y}$. These symbols are then used to produce the reproduction sequence $\hat{x_1}, \hat{x_2}, ...$ in the alphabet $\hat{\mathcal{X}}$. The rate of the source coding is defined as the ratio of the length of the output sequence to the input sequence (normalized by the size of alphabets). The decoding scheme is called $t-$local if, for any $i = 1, 2, ...$, the reproduced symbol, $\hat{x_i}$ only depends on t of $y_1, y_2, ...$ (t is called the number of queries). In conventional source coding, $\hat{X^n}$ depends on $\hat{Y^k}$ in an arbitrary manner. In this work, we characterize the source coding rate for the setting in which the decoder is constrained to ask only t queries to reproduce any reproduction source coordinate , i.e., $\hat{x_i}$. Figure 3-1 demonstrate the problem formulation. As it shows, any y_i is an arbitrary function of $x_1, x_2, ...$ and any $\hat{x_i}$ is a function of only t of $y_1, y_2, ...$

This problem appears in many applications in distributed data management. For instance, assume a given source is stored in some storage cells. Since writing on data storage cells is generally costly, we use source coding to decrease the number of cells used. If we wish to recover a part of the original source (in our case one coordinate of the source sequence) we may need to read the entire encoded data on all data storage cells. Therefore, we need to query all the cells and then decode the coordinate of the source that we wanted to recover. However, we know that reading from the storage cells is generally costly, so we wish to read as few cells as possible. Clearly, there is a trade off between the number

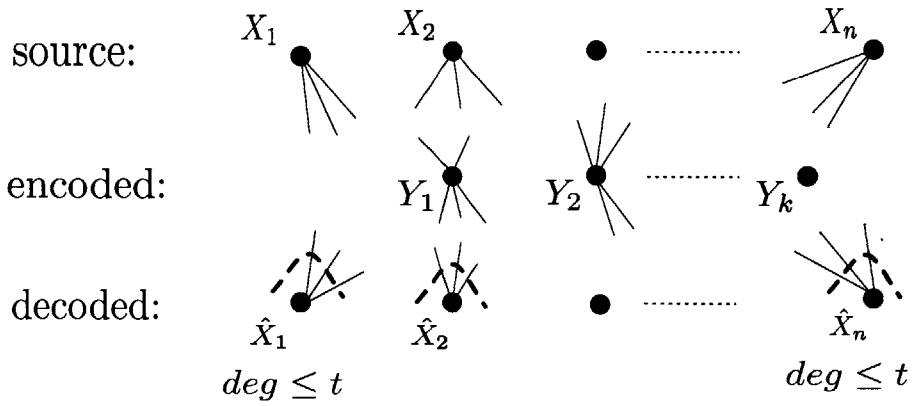

Figure 3-1: Locally Decodable Source Coding

of used storage cells to store the entire original source (rate) and the number of cells we need to access in order to recover a coordinate of the original source sequence (locality, t). Characterizing this trade-off is one of our goals.

In another example, assume that we encode a source and then store it on some data storage cells. We want to reveal the information about one coordinate of the source to some party, but, we do not want to reveal the information about the entire source symbols. If we use a conventional source coding, we may have to reveal all the encoded data. Thus, a honest but curious party may have access to the entire original source sequence. On the other hand, in LDSC we provide only a small part of the encoded data, so the party can only recover the desired part of original source symbols without capability of extracting the other symbols.

3.1 Lossless Locally Decodable Source Coding

In this section, we study lossless source coding in the presence of local decoders. We formally define LDSC and study its rate.

3.1.1 LDSC

A local decoder is a decoder that only asks a few number of queries to decode any coordinate of the source sequence. The number of queries is denoted by t. The formal definition is

as follows. A lossless LDSC is defined as a pair consisting of an encoder, f, and a decoder, g, where, $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$. The decoder is called local if each coordinate of the output is affected by a bounded number of input coordinates. Formally, Let g_a, for $a \in \{1, ..., n\}$, be the $a-$th component of the decoding function. Assume g_a depends on Y^k only through the vector $Y^{\mathcal{N}_a^X} = \{Y_j \ : \ j \in \mathcal{N}_a^X\}$ for some $\mathcal{N}_a^X \subset \{1, ..., k\}$, meaning that

$$\text{For any } y^k \text{ and } y'^k, g_a(y^k) = g_a(y'^k) \text{ if } y^{\mathcal{N}_a^X} = y'^{\mathcal{N}_a^X}.$$

For any given t, a decoder is called $t-$local if $|\mathcal{N}_a^X| \leq t$ for any $a \in \{1, ..., n\}$.

Definition 4. *An $(n, k, t, \epsilon)-$LDSC is a pair consisting of an encoder $f : X^n \mapsto \{0,1\}^k$ and a $t-$local decoder $g : \{0,1\}^k \mapsto X^n$, such that*

$$\mathbb{P}[g(f(X^n)) \neq X^n] \leq \epsilon. \tag{3.1}$$

Using a similar notation to that of [21], let

$$k_{ld}^*(n, \epsilon, t) \triangleq \min\{k : \ \exists \ (n, k, \epsilon, t) - \textit{LDSC}\}. \tag{3.2}$$

Where the subscript ld stands for local decoder. The best rate of local code for a given n, t and ϵ, is given by

$$R_{ld}(n, \epsilon, t) \triangleq \frac{k_{ld}^*(n, \epsilon, t)}{n}. \tag{3.3}$$

Also let

$$R_{ld}(\epsilon, t) \triangleq \limsup_{n \to \infty} R_{ld}(n, \epsilon, t). \tag{3.4}$$

We define the rate as

$$R_{ld}(t) \triangleq \lim_{\epsilon \to 0} R_{ld}(\epsilon, t). \tag{3.5}$$

Lemma 1. *Given a $(n, k, \epsilon, t) - $ LDSC with randomized encoder and decoder, there exists*

an (n, k, ϵ, t) *– LDSC code with deterministic encoder and decoder.*

Proof: Let M and N be two random variables, and consider randomized encoder and decoder $f(M)$ and $g(N)$, respectively. Equation (3.1) then becomes

$$\mathbb{P}[g(f(X^n, M), N) \neq X^n] = \mathbb{E}[\mathbb{P}[g(f(X^n, M), N) \neq X^n]|M, N] \leq \epsilon.$$

Since the probability in equation (3.1) is less than or equal to ϵ, then there exist m, n such that

$$\mathbb{P}[g(f(X^n, M), N) \neq X_i|M = m, N = n] \leq \epsilon,$$

implying that $f(m)$ and $g(n)$ are our desired sort of encoder and decoder, respectively, and the proof is complete. □

Note 3. *Using Lemma 1, in the rest of the text, we assume the encoder and decoder are deterministic.*

3.1.2 Average LDSC

Instead of assuming the number of queries to recover any X_i is bounded, consider the case where the average number of queries asked to recover all the X_i s is bounded. We term this *Average Locally Decodable Source Coding (ALDSC)*. The formal definition is the following.

Definition 5. *For any given t, a decoder is called* $t-$*average local if* $\frac{1}{n}\sum_{a=1}^{n} |\mathcal{N}_a^X| \leq t$. *An* $(n, k, t, \epsilon)-$*ALDSC is a pair consisting of an encoder* $f : X^n \mapsto \{0, 1\}^k$ *and a* $t-$ *average local decoder,* $g : \{0, 1\}^k \mapsto X^n$, *such that*

$$\mathbb{P}[g(f(X^n)) \neq X^n] \leq \epsilon.$$

Similarly, define

$$k^*_{ald}(n, \epsilon, t) \triangleq \min\{k : \ \exists \text{ an}(n, k, \epsilon, t) - \text{ALDSC}\},$$

where the subscript, ald, stands for average local decoder. The best rate of a ALDSC for a given n, t and ϵ, is given by

$$R_{ald}(n, \epsilon, t) \triangleq \frac{k^*_{ald}(n, \epsilon, t)}{n}.$$

Also, define

$$R_{ald}(\epsilon, t) \triangleq \limsup_{n \to \infty} R_{ald}(n, \epsilon, t).$$

We define the rate as

$$R_{ald}(t) \triangleq \lim_{\epsilon \to 0} R_{ald}(\epsilon, t).$$

The following result establishes a relation between $R_{ld}(t)$ and $R_{ald}(t)$.

Proposition 1. *For any* $0 \leq \lambda \leq 1$, *we have* $k^*_{ld}(n, \epsilon, t) \geq k^*_{ald}(\lambda n, k, \epsilon, \frac{t}{1-\lambda})$.

Proof: The proof follows from sorting the number queries for all of the source symbols and then selecting the first λ fraction of them. Let (n, k, t, ϵ) be an ALDSC. Let t_i denote $|\mathcal{N}_i|$ for $1 \leq i \leq n$. Without loss of generality, assume $t_1 \leq t_2 \leq \ldots \leq t_n$. We have $\frac{1}{n}\sum_{i=1}^{n} t_i \leq t$ (for the sake of presentation, without loss of generality, we drop ceiling and floor). Therefore, we get $t_{\lambda n} \leq \frac{t}{1-\lambda}$. The corresponding decoder and encoder introduces a $(\lambda n, k, \frac{t}{1-\lambda}, \epsilon) -$ LDSC. Hence,

$$\{\, k \mid \exists (n, k, \epsilon, t) - \text{ ALDSC}\} \subseteq \{\, k \mid \exists (\lambda n, k, \epsilon, \frac{t}{1-\lambda}) - \text{LDSC}\}$$
$$\Rightarrow k^*_{ald}(n, \epsilon, t) \geq k^*_{ld}(\lambda n, k, \epsilon, \frac{t}{1-\lambda}).$$

The proof is complete. □

Corollary 1. *For any* $0 \leq \lambda \leq 1$, *we have* $\lambda R_{ld}(\frac{t}{1-\lambda}) \leq R_{ald}(t)$.

Proof: Using Proposition 1,

$$\lambda R_{ld}(\frac{t}{1-\lambda}) = \lambda \frac{k^*_{ld}(\lambda n, \epsilon, \frac{t}{1-\lambda})}{\lambda n} \leq \frac{k^*_{ald}(n, k, \epsilon, t)}{n} = R_{ald}(t),$$

which concludes the proof. □

This corollary states that, once we have a lower bound on the rate of LDSC, we obtain a lower bound on the rate of ALDSC. Implying that we would not gain much, by using ALDSC instead of LDSC. We shall quantify this notion in next sections.

3.1.3 Linear Encoder

Source coding with a linear encoder and local decoder is the subject of this section. We focus on binary sources where $\mathcal{X} = \{0, 1\}$. We show that, for a linear encoder, the rate of LDSC is one instead of the entropy rate, implying no compression is possible. Before proving this fact, we introduce the influence matrix of an encoder.

Definition 6. *Let $f : \{0,1\}^n \mapsto \{0,1\}$ be a Boolean function defined on $\{0,1\}^n$. The influence of X_i (the ith component of f) on f is defined as*

$$Inf_i(f) \triangleq \mathbb{P}[f(x + e_i) \neq f(x)]. \tag{3.6}$$

An encoder $f : X^n \mapsto \{0,1\}^k$ can be treated as a collection of k Boolean functions on $\{0,1\}^n$. If we denote the jth function by f_j, then $f(x) = (f_1(x), ..., f_k(x))$. The influence matrix of the encoder $f : X^n \mapsto \{0,1\}^k$ is an $n \times k$, matrix, A, defined as $A_{ij} = Inf_i(f_j)$. Note that if a function is linear, then the influence matrix in the same as the generating matrix.

We illustrate a relation between a function Influence and its ability to decode a particular bit by an example.

Example 1. *In this example, we show that two functions f_1 and f_2 may recover X_1, while having almost zero information about X_1. Let $f_1 = X_1 + ... + X_n$, and $f_2 = X_2 + ... + X_n$, where the summation is in $\mathbb{F}_2$. Here, we have perfect recovery by $X_1 = f_1 + f_2$. We have $Inf_1(f_1) = 1$ while $I(X_1; f_1) = h(\frac{1}{2}(1 - (1 - 2p)^n) - h(\frac{1}{2}(1 - (1 - 2p)^{n-1})$ which is very close to zero for large n. On the other hand, $Inf_1(f_2) = 0$ and also $I(X_1; f_2) = 0$. Therefore, influence of X_1 may be either 0 or 1, while the mutual information remains zero.*

Next, we prove a converse bound on the a rate of LDSC with linear encoding. In order to prove the theorem, we use the following lemma.

Lemma 2. *Let $\mathbb{F}_2^n$ be a vector space over $\mathbb{F}_2$. Let $\mathbb{P}_X = Bern(p)$ and define a probability measure over $\mathbb{F}_2^n$ according to $\mathbb{P}_X^n$. If U is a k- dimensional sub-space of $\mathbb{F}_2^n$, we have*

$$(\max\{p, 1-p\})^{n-k} \geq \mathbb{P}[U] \geq (\min\{p, 1-p\})^{n-k}.$$

Proof: We first prove the lower bound. Define $E = \{v \in V | H(v) = 1\}$, where $H(v)$ denotes the hamming distance of v. Since the dimension of U is k, there is E' a subset of E with $n-k$ elements such that

$$U \oplus U' = \mathbb{F}_2^n$$
$$U \cap U' = \{0\},$$

where $U' = span\{E'\}$ and $\oplus$ denotes the direct sum of two sub spaces. For each $u' \in U'$ define $U_{u'} = U + u'$. It is clear that $U_{u'_1} \cap U_{u'_2} = \emptyset$ for $u'_1 \neq u'_2$. Next, we shall bound $P(U_{u'})$. Suppose $H(u') = r$, then we have:

$$\begin{aligned}\mathbb{P}[U_{u'}] &= \sum_{u \in U_{u'}} \mathbb{P}[u] = \sum_{u \in U} \mathbb{P}[u + u'] \\ &\geq \sum_{u \in U} \mathbb{P}[u] \left(\frac{\min\{p, 1-p\}}{\max\{p, 1-p\}}\right)^r = \mathbb{P}[U] \left(\frac{\min\{p, 1-p\}}{\max\{p, 1-p\}}\right)^r\end{aligned}$$

Since $U_{u'}$s are disjoint and for each $u' \in U'$, we have $H(u') \leq n-k$ and the following equation holds

$$\begin{aligned}1 = \mathbb{P}[\mathbb{F}_2^n] &= \mathbb{P}[\cup_{u' \in U'} U_{u'}] = \sum_{u' \in U'} \mathbb{P}[U_{u'}] \\ \leq \sum_{u' \in U'} \mathbb{P}[U] \left(\frac{\max\{p, 1-p\}}{\min\{p, 1-p\}}\right)^r &= \mathbb{P}[U] \sum_{r=0}^{n-k} \binom{n-k}{r} \left(\frac{\max\{p, 1-p\}}{\min\{p, 1-p\}}\right)^r \\ = \mathbb{P}[U] \left(1 + \frac{\max\{p, 1-p\}}{\min\{p, 1-p\}}\right)^{n-k} &= \mathbb{P}[U] \left(\frac{1}{\min\{p, 1-p\}}\right)^{n-k} \qquad (3.7)\end{aligned}$$

Changing the second line of (3.7), the upper bound is proved similarly. □

Theorem 7. *Assume X has a $Bern(p)$ distribution and (n, k, ϵ, t) is a LDSC for this source with a linear encoder. If $\epsilon < (\min\{p, 1-p\})^t$, then $k \geq n$.*

Proof: Let G be the corresponding influence matrix of the encoder. The encoding is then as following

$$x \mapsto xG.$$

Since the encoder is linear, the entries of the influence matrix are either 0 or 1, meaning $G \in \mathbb{F}_2^{n\times k}$. G is a mapping from $\{0,1\}^n$ to $\{0,1\}^k$. Without loss of generality, assume that X_1 is recovered by $Y_1, ..., Y_t$ and the decoder maps 0^t to $\hat{X}_1 = 0$. Consider the induced linear mapping $\pi : X^n \to Y^t$, we have $\dim(ker(G)) \geq n - t$. Note that $0^n \in ker(\pi)$. If there exists $x^n \in ker(\pi)$ such that $x_1 = 1$, then half of the vectors in $ker(\pi)$ have $x_1 = 0$ and half of them have $x_1 = 1$. Since the decoder maps 0^t to $\hat{X}_1 = 0$, then the vectors in $ker(\pi)$ with $x_1 = 1$ are erroneous . Eliminate the first coordinate and consider all the vectors in $ker(\pi)$ such that $x_1 = 1$, they will form a subspace of dimension at least $n-t-1$ is a paces of dimension $n-1$. Therefore using Lemma 2

$$\mathbb{P}[\hat{X}^n \neq X^n] \geq \mathbb{P}[\hat{X}_1 \neq X_1] \geq \mathbb{P}[S] \geq (\min\{p, 1-p\})^{n-1-(n-t-1)} = (\min\{p, 1-p\})^t.$$

Which is a contradiction.

Hence, for any $x^n \in ker(\pi)$, $x_1 = 0$. This means that, if we look at the sub-matrix of G of dimension $n \times t$ consisting of the first t columns, the first row is not in the span of the rest of rows. This implies that, in the matrix G, the first row is not in the span of the rest of rows. If we apply the same argument for any $\hat{X}_i$, we conclude that the rows of the matrix G are independent, resulting in $k \geq n$. □

Corollary 2. *Using linear encoder, for any t, we have $R_{ld}(t) = 1$. Moreover, by using Proposition 1, we have $R_{ald}(t) = 1$.*

Proof: It directly follows from Theorem 7 and Corollary 1.

In contrast with Theorem 4 about LESC, LDSC does not achieve the fundamental limit on

the rate, which is entropy, meaning that we can not compress data, expecting to recover it locally.

Note that, Theorem 7 also proves that using a linear encoder and local decoder, we can only have lossless (with zero error) codes.

3.1.4 Linear Decoder

In this section, we assume that the decoder is local and linear. We show that, for a linear decoder, even without locality constraint, the rate of compression is 1. This implies that, if the decoder is linear, then no compression is possible.

Theorem 8. *Let X have Bern (p) distribution. Assume the decoder is linear. We have*

$$k^*_{sc}(n,\epsilon) \geq n - \frac{\log(1-\epsilon)}{\log\left(\max\{p, 1-p\}\right)} \tag{3.8}$$

Proof:

Consider a (n, k, ϵ)- SC. Assume $e_1, ..., e_k$ are the canonical basis of $\{0,1\}^k$. Decoder can only recover $Span\{g(e_1), ..., g(e_k)\}$ and we have error for all other inputs. Thus, using Lemma 2

$$P[g(f(X^n)) \neq X^n] \geq 1 - \mathbb{P}[Span\{g(e_1), ..., g(e_k)\}] \geq 1 - (\max\{p, 1-p\})^{n-k}$$

we also know $P[g(f(X^n)) \neq X^n] \leq \epsilon$. Therefore,

$$k \geq n - \frac{\log(1-\epsilon)}{\log \max\{p, 1-p\}}.$$

Taking the minimum over all choices of codes, we get

$$k^*_{sc}(n,\epsilon) \geq n - \frac{\log(1-\epsilon)}{\log\left(\max\{p, 1-p\}\right)},$$

and the proof is complete. □

Let X be a Bern(p) source and $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$ be the general encoder and linear decoder. Theorem 8 shows that that $R_{sc} = 1$. In other words, if we

employ linear decoding, the rate of compression is always one, instead of entropy rate. Therefore, for any $t-$local decoder we have

$$R_{ld}(t) = 1.$$

Moreover, using Corollary 1 for a linear decoder, we obtain

$$R_{ald}(t) = 1.$$

3.1.5 General Encoder-Decoder

In previous sections, we showed that assuming a linear encoder or linear encoder the rate of LDSC is 1. We study the same problem for a general encoder-decoder. First, by using an example, we show the existence of good non-linear encoders with local decoder.

Example 2. *Consider the following f:*

$$f_0 = X_1 \text{ or } X_2 \text{ or } ... \text{ or } X_n, \quad f_i = X_i \text{ or } (X_1 \text{ or } ...X_{i-1} \text{ or } X_{i+1} \text{ or } ... \text{ or } X_n \ + \ 1).$$

The recovery is as

$$\hat{X}_i = f_0 \text{ and } f_i.$$

In this example we have zero error and $k = n + 1$. This shows the existence of completely non-linear codes.

The following theorem focuses on the special case of $t = 2$. Without loss of generality, we assume $p \leq \frac{1}{2}$.

Theorem 9. *Let X be a Bern (p) source and $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$ be a general encoder and t-local decoder, respectively. Also assume a (n, k, ϵ, t)-LDSC for this source. For $t = 2$, if $\epsilon < p^2$, then $k \geq n$.*

Proof: We prove this by contradiction. For the sake of contradiction, assume $n > k$, we show that $\epsilon > p^2$, which is a contradiction.

The claim is that if the code can recover X_1^{k+1} $(i.i.d$ Bern$(p))$ with a local map from Y_1^k

on a set with probability $p(k)$, then $p(k) < 1 - p^2$. Note that this is enough to prove the theorem because $\epsilon = 1 - p(k) \geq p^2$

By induction on k, we show that $p(k) < 1 - p^2$. It can be shown that $p(1) \leq 1 - p^2$. Let $p(k-1) \leq 1 - p^2$. Assume X_1 is recovered by Y_1 and Y_2. Without loss of generality, assume that $g_1(0,0) = 0$, where for any $1 \leq i \leq n$, g_i is a mapping, with two inputs, that reproduces X_i. Here are the all possible cases:

1. $g_1(0,1) = 0$. In this case, if we consider the induced map from Y_2^k to X_2^{k+1} by replacing 0 with Y_1 in all the mappings that use Y_1 as one of their inputs, we end up with a local mapping on a set with maximum probability of $p(k-1)$. Similarly, since $g_1(1,1) = g_1(1,0) = 1$, if we replace 1 with Y_1, we get another local mapping on a set with maximum probability $p(k-1)$. Therefore, $p(k) \leq p.p(k-1) + 1 - p.p(k-1) = p(k-1) \leq 1 - p^2$.

2. $g_1(1,0) = 0$. In this case, replace 0 with Y_2 and construct a mapping from Y_1, Y_2^k to X_2^{k+1}. Similarly, it can be shown that $p(k) \leq 1 - p^2$.

3. $g_1(1,1) = 0$. In this case, replace Y_1 by Y_2 in all the mappings that use Y_1 as one of their inputs. Similarly, we obtain $p(k) \leq 1 - p^2$.

4. $g_1(1,0) = g_1(0,1) = g_1(1,1) = 1$. In this case, $g_1(Y_1, Y_2) = \bar{Y_1}.\bar{Y_2}$. This case requires more details which will follows.

We call $Y_1.Y_2$, $\bar{Y_1}.Y_2$, $Y_1.\bar{Y_2}$, and $\bar{Y_1}.\bar{Y_2}$ a product form. The discussion which we provided before shows that if only one of the $k+1$ mappings is not of the product form, then the above argument proves $p(k) \leq 1 - p^2$. Now, assume all the mappings are in the product form. If Y_i is once used as $X_{i_1} = Y_i.Y_j$ and once as $X_{i_2} = \bar{Y_i}.Y_k$, then $X_1 = X_2 = 1$ cannot be recovered. Implying $p(k) \leq 1 - \bar{p}^2 \leq 1 - p^2$ ($p < \frac{1}{2}$). Therefore, without loss of generality, we assume that all the mapping functions are of the form $Y_i.Y_j$.

There are $k+1$ mappings, so the number of arguments of all these mappings is $2(k+1)$. Also, note that there are k Y_i's. Thus, three mappings can be found as $X_{i_1} = Y_iY_j$, $X_{i_2} = Y_iY_k$, and $X_{i_3} = Y_jY_l$. If $X_{i_1} = 1$, then we can find a mapping from Y_1^{i-1}, Y_{i+1}^k to $X_1^{i_1}, X_{i_1+1}^{k+1}$ on a set with maximum probability $p(k-1)$. Also note that $X_{i_1} = 0$,

$X_{i_2} = X_{i_3} = 1$ is not possible to recover. Therefore, $p(k) \leq p.p(k-1) + \bar{p}.p^2 \leq 1 - p^2$. The proof is complete for $t = 2$. □

Corollary 3. *Let X be a Bern(p) source. For $t = 2$, $R_{ld}(t) = 1$.*

Proof: It directly follows from Theorem 9.

Also, for ALDSC we obtain the following.

Corollary 4. *Let X be a Bernoulli (p) source. For $t = 2$, $R_{ald}(t) = 1$.*

Proof: It follows from Theorem 9 and Corollary 1. □

Considering the bipartite graph describing the relation between input and output of a general decoder, one way to generalize this proof to $t \geq 3$ is to find a small number of encoded coordinates that are connected to a small number of recovered source coordinates. The following section discusses this idea. We have the following conjecture for any given $t > 2$: Let X be a Bern(p) source and $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$ be a general encoder and t-local decoder. For any t, we have $R_{ld}(t) = 1$. As a result of Corollary 1, $R_{ald}(t) = 1$ follows.

3.1.6 On the Bounded Degree Bipartite Graph

In previous section, in the case of $t = 2$, we found two $\hat{X}_i$'s which are connected to only two of Y_j's. Then we considered all the possibilities for this sub-mapping and then did induction to prove Theorem 9. If we can find a subset of the $\hat{X}_i$'s with size s which are connected to a small number of Y_i's such as $s + O(1)$, then we would follow the same approach as of the proof of Theorem 9. The following result rules out this approach.

Let $G(U, V, E)$ be a bipartite graph, where U and V are two set of nodes connected with the edges from E.

Definition 7. *We call a bipartite graph $G(U, V, E)$ a $t-$ regular bipartite graph if any node in V has a degree equal to t. Moreover, let the set $G(n, m, t) = \{G(U, V, E)\ t-$ regular such that $|U| = m,\ \ |V| = n\}$.*

For a given subset of the nodes V such as I, let $N(I)$ denote the neighbors of the nodes in I.

Definition 8. *For any given l, let*

$$N(l,n,m,t) = \max_{G \in G(n,m,t)} \min_{I:|I|=l} N(I).$$

In order to obtain insight about $N(l,n,m,t)$, consider the following example.

Example 3. *If $m \geq nt$, then $N(l,n,m,t) = lt$ and the graph that achieves this is the one for which all nodes in V are connected to disjoint set of t nodes of U. If m is very small relative to n, then $N(l,n,m,t)$ is below lt.*

We are interested in the case where $m = n-1$. The question we try to answer is the following: for a given t and l, what is $N(l,n,n-1,t)$, for large enough n. The following theorem shows that the number of neighbors of a set of size l is growing at least linearly with l.

Theorem 10. *For any given t, δ and l, there exists $n_0(t,l,\delta)$ such that, for $n \geq n_0(t,l,\delta)$*

$$N(l,n,n-1,t) \geq (t-1-\delta)l.$$

Therefore, $N(l,n,n-1,t)$ grows not slower than linearly with l, for a given t.

Proof: Let the set $X = \{1,...,n-1\}$ denote the nodes in U. Let $A_1,...,A_k$ be k sub-sets of X, each of them with t elements corresponding to the neighbors of the v in X, where $v \in V$. Also, assume that the union of any l of these subsets has at least $(t-1-\delta)l$ many elements. We must show that, for large n, there exists a collection of size n of these subsets, i.e. k can be equal to or larger than n. We use a probabilistic approach to show this. Assume that the subsets are chosen uniformly from X. Let $\mathbf{1}\{|\cup_{j=1}^{l} A_{i_j}| < (t-1-\delta)l\}$ be the indication function which is one if its argument holds and is zero otherwise. Now, consider the following expectation for a given k,n,t,l:

$$\mathbb{E}[\sum_{i_1,...,i_l} \mathbf{1}\{|\cup_{j=1}^{l} A_{i_j}| < (t-1-\delta)l\}].$$

If this quantity is less than one, then for the given k,n,t,l there exists a bipartite graph in $G(k,n-1,t)$ such that, for any subset of l nodes of V, the union of their neighbors has

more than or equal to $(t-1-\delta)l$ node. Thus, in order to complete the proof, we show that there exists $n_0(t,l,\delta)$ for which, if $n \geq n_0(t,l,\delta)$, then the expectation becomes less than one for $k=n$.

We have

$$\mathbb{E}\left[\sum_{i_1,\ldots,i_l} \mathbf{1}\{|\cup_{j=1}^{l} A_{i_j}| < (t-1-\delta)l\}\right] = \binom{k}{l}\mathbb{P}\left[|\cup_{i=1}^{l} A_i| < (t-1-\delta)l\right].$$

Let $N(t,l)$ denote the number of different ways in which one can choose l subsets of a set with $(t-1-\delta)l$ elements, each of them having t elements. We obtain

$$\mathbb{P}\left[|\cup_{i=1}^{l} A_i| < (t-1-\delta)l\right] = \frac{\binom{n-1}{(t-1-\delta)l}}{\binom{n-1}{t}^l} N(t,l).$$

Hence,

$$\mathbb{E}\left[\sum_{i_1,\ldots,i_l} \mathbf{1}\{|\cup_{j=1}^{l} A_{i_j}| < (t-1-\delta)l\}\right] = \binom{k}{l}\frac{\binom{n-1}{(t-1-\delta)l}}{\binom{n-1}{t}^l} N(t,l)$$
$$\leq k^l \frac{(n-1)^{(t-1-\delta)l}}{\left(\frac{(n-1-t)^t}{t!}\right)^l} N(t,l) \leq \left(\frac{n(n-1)^{(t-1-\delta)}}{(n-1-t)^t}\right)^l (t!)^l N(t,l)$$

Comparing the exponent of n in the numerator and denumerator, there exists $n_0(t,l,\delta)$ (some number which depends on t, l, and δ) with the property given in the theorem. □

3.1.7 Scaling Number of Queries

As we mentioned before, the number of queries, t can be a growing function of n. In the conventional source coding (not necessarily local) $t(n)$ is a linear function of n. Therefore, the regime of our interest is the one where $t(n)$ grows sub linearly with n. In order to establish the results on LDSC with scaling number of queries, we need the following results on the error exponent of source coding to establish an achievability bound on the rate. This approach is motivated by the achievability bound given in reference [15].

Theorem 11. *([6]) For a discrete memoryless source with probability measure $\mathbb{P}_X$ and a*

source encoding with rate R, we have:

For any $\epsilon > 0, \exists\, K_\epsilon$ such that for any $n \geq 0$ there exists an encoding-decoding pair f_n and g_n such that

$$\mathbb{P}[g_n(f_n(X^n)) \neq X^n] \leq K_\epsilon 2^{-n(E_b^*(R)-\epsilon)}. \tag{3.9}$$

Where

$$E_b^*(R) = \min_{Q:H(Q)\geq R} D(Q||P).$$

Moreover, this bound is asymptotically tight.

Now, consider the following construction of an encoder-decoder for a sequence of length n of the source Bern(p):

Let the rate R be equal to $(1+\delta)H(X)$. Let X^n be a sequence of source symbols. Divide this sequence into blocks of length $t(n)$ and apply the encoder-decoder pair found by Theorem 11 to each block separately. Form an encoder-decoder for X^n by cocatenating these $\frac{n}{t(n)}$ (for the sake of presentation, without loss of generality, we drop ceiling and floor in this analysis) pairs of encoder-decoder. We now analyze the error of the concatenated source coding. Using a union bound we obtain

$$\mathbb{P}[\hat{X}^n \neq X^n] = \mathbb{P}[\cup_{i=1}^{n/t(n)}\{\hat{X}^{it(n)}_{(i-1)t(n)+1} \neq X^{it(n)}_{(i-1)t(n)+1}\}] \leq \frac{n}{t(n)}\mathbb{P}[\hat{X}^{t(n)} \neq X^{t(n)}].$$

Using, (3.9) for any ϵ, we obtain

$$\mathbb{P}[\hat{X}^n \neq X^n] \leq \frac{n}{t(n)} K_\epsilon 2^{-t(n)(E_b^*(R)-\epsilon)}.$$

Since $R > H(X)$, $E_b^*(R) = \Delta > 0$. Thus, we have

$$\mathbb{P}[\hat{X}^n \neq X^n] \leq \frac{n}{t(n)} K_\epsilon 2^{-t(n)(\Delta-\epsilon)}.$$

Choose $\epsilon < \Delta$ and denote $\Delta - \epsilon$ by Γ. We want to see for which choices of $t(n)$, this bound goes to zero. This happens when

$$\lim_{n\to\infty} \Gamma t(n) - \log \frac{n}{t(n)} = \infty$$

resulting in

$$t(n) > C \log n,$$

for some constant C. Therefore, we have the following result.

Proposition 2. *Let X be a Bern (p) source and $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$ be an encoder and $t(n)$-local decoder. For any ϵ and $R > H(X)$, there exists a C and n_0 such that for $n > n_0$, there exist a $(n, nR, C \log n, \epsilon)$-LDSC.*

This theorem states that with number of queries $(\log n)$ that is small relative to n, the rate of LDSC approaches the optimal rate $h(p)$.

The same results also holds for linear codes. In particular, we use the following results.

Theorem 12. *([5]) For a given source with probability measure $\mathbb{P}_X$, n, and k there exists a linear encoder, $f : \mathcal{X}^n \to \mathcal{X}^k$ and and a decoder such that*

$$\mathbb{P}[\hat{X}^n \neq X^n] \leq 2^{-n\left((R-H(P))-2|\mathcal{X}|^2 \log \frac{n+1}{n}\right)}.$$

Using this theorem we get the following result for linear encoders and local decoders.

Corollary 5. *Let X be a Bern(p) source and let $f : X^n \mapsto \{0,1\}^k$ and $g : \{0,1\}^k \mapsto X^n$ be a linear encoder and $t(n)$-local decoder, respectively. For any $\epsilon > 0$ and $R > H(X)$, there exists a C and n_0 such that for $n > n_0$ there exist a $(n, nR, C \log n, \epsilon)$-LDSC.*

3.2 Lossy Locally Decodable Source Coding

In this section, *Locally Decodable Lossy source Coding (LDLSC)* is studied. First, we formally define the problem.

Consider a separable distortion metric defined as $d(X^n, \hat{X}^n) = \frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)$.

Definition 9. *A (n, k, d, t)-LDLSC is a pair consisting of an encoder $f : X^n \mapsto Y^k$ and a decoder $g : Y^k \mapsto X^n$, where the decoder is t-local. The average distortion is bounded, $\mathbb{E}[d(X^n, g(f(X^n)))] \leq d$.*

Let

$$k^*_{ld}(n, d, t) \triangleq \min\{k \text{ such that } \exists (n, k, d, t) - \text{ LDLSC }\}, \tag{3.10}$$

where the subscript ,ld, stands for local decoder, also

$$R_{ld}(d,t) \triangleq \limsup_{n\to\infty} \frac{k^*_{ld}(n,d,t)}{n}. \tag{3.11}$$

Note 4. *We assume a binary source, $\mathbb{F} = \mathbb{F}_2$ with $d(X,\hat{X}) = \mathbf{1}\{X \neq \hat{X}\}$. In this case, we have*

$$\mathbb{E}[d(X^n, g(f(X^n)))] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}[X_i \neq \hat{X}_i] \leq d,$$

which is the same as assuming the bit error rate is bounded (relative to the block error rate in the definition of LDSC).

3.2.1 Scaling number of queries

In this section, we consider the scaling of number of queries. Therefore, let $t(n)$ shows the number of queries. The following is an upper bound on the rate for scaling number of queries.

Theorem 13. *For a Bernoulli(p) source, a distortion level d, and any increasing number of queries $t(n)$*

$$R_{ld}(n,d,t(n)) \leq h(p) - h(d) + \frac{\log t(n)}{t(n)} + o(\frac{\log t(n)}{t(n)}) \tag{3.12}$$

Proof: Recall the finite block length results on source coding ([27]) thereof for a Bern(p) source, and distortion level d, there exists a code such that

$$R_{lsc}(n,d) \leq h(p) - h(d) + \frac{\log n}{n} + o(\frac{\log n}{n}). \tag{3.13}$$

Now, divide the sequence X^n into $\frac{n}{t(n)}$ blocks of length $t(n)$. Apply the encoder-decoder obtained from (3.13) to each block. Concatenate these $\frac{n}{t(n)}$ pairs to obtain an encoder-decoder for X^n. The average distortion of the overall code is also bounded by d, and its rate is bounded by

$$\left(\frac{n}{t(n)}\left(t(n)\left(h(p) - h(d) + \frac{\log t(n)}{t(n)} + o(\frac{\log t(n)}{t(n)})\right)\right)\right)/n,$$

which shows the theorem. □

Theorem 13 shows that, for any number of queries $t(n)$, if $\lim_{n\to\infty} t(n) = \infty$, then there exists $(n, k, D, t(n))$-LDLSC such that $k \leq n(h(p) - h(D)) + o(n)$. Note that without loss of generality, we dropped ceiling and floor in the analysis.

Corollary 6. *For the special case $t(n) = t\log n$, we have*

$$R_{ld}(n, d, t\log n) \leq h(p) - h(d) + \frac{\log t \log n}{t\log n} + o(\frac{\log t \log n}{t\log n}) \tag{3.14}$$

Proof: The result of Theorem 13, for $t(n) = t\log n$. □

Patrascu ([20]) studies the problem of storage of bits with local recovery of bits (with the same definition of locality we use here). Those results are based on a generic transformation of augmented B-trees to succinct data structures. He shows that

Theorem 14 ([20]). *Consider an array of length n from alphabet $\mathcal{X}$. We can represent this array with a number of bits*

$$O(|\mathcal{X}|\log n) + n\tilde{H} + \frac{n}{(\frac{\log n}{t})^t} + O(n^{3/4}\log n^{3/4}), \tag{3.15}$$

supporting single bit recovery in $t\log n$ queries, where $\tilde{H}$ denotes the empirical entropy (the entropy of the empirical distribution). Moreover, we can represent a binary sequence of length u, with n ones, using

$$\log\binom{u}{n} + \frac{u}{(\frac{\log u}{t})^t} + O(u^{3/4}\log u^{3/4}) \tag{3.16}$$

bits, for which a decoder exists querying only $t\log u$ bits. Note that the encoder/decoder knows n and u beforehand.

Reference [20] essentially studies the case of $t(n) = t\log n$. We now compare the bound given in corollary 6 with the bound suggested by Theorem 14.

Using Theorem 14 and identity $\log\binom{n}{pn} = nh(p) + O(\log n)$, for any d, we obtain

$$R_l(n, d, t\log n) \leq h(p) + O(\frac{\log n}{n}) + \frac{1}{(\frac{\log n}{t})^t} + O(\frac{\log n^{3/4}}{n^{1/4}}). \tag{3.17}$$

It is clear that, for any fixed d, the bound given by (3.14) is asymptotically tighter than (3.17). The comparison is illustrated in Figures 3-2, 3-3. The curves are approximations of the bounds, because we omit the last term of each of the bounds. As it can be seen in the figures, for any large enough n, the bound (3.14) is tighter than (3.17); for small n and small distortion level, d, the bound given in (3.17) can be tighter. One may consider the

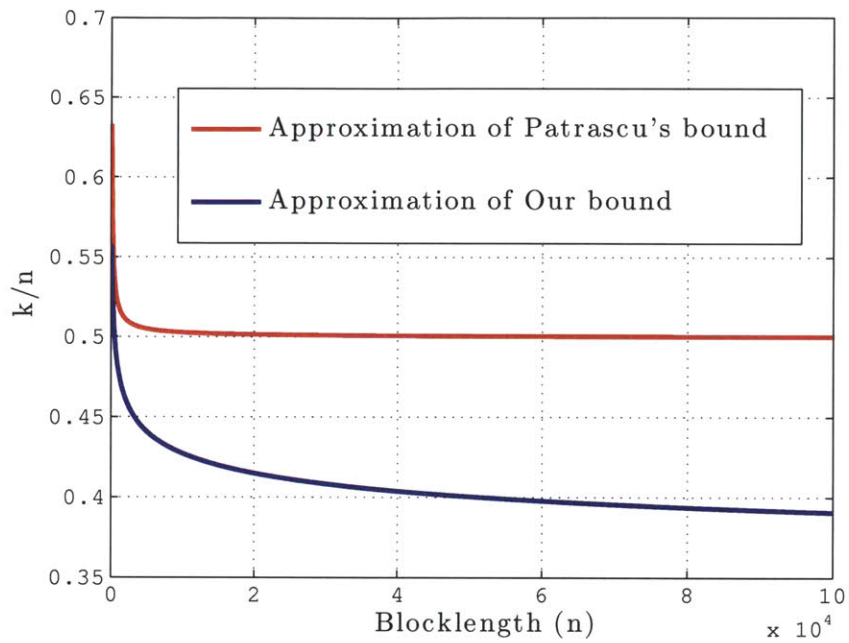

Figure 3-2: Comparison of approximations of upper bounds, $p = .11$, $d = 1/15$

case where d goes to zero as n goes to infinity. Assume both bounds hold for this case as well. We show that if $d(n) = O(\frac{1}{\log n})$, then (3.17) is tighter than (3.14). We omit the last term in both bounds and assume $t = 1$. In order to show that, (3.17) is tighter than (3.14), we need to prove

$$h(p) + O(\frac{\log n}{n}) + \frac{1}{\log n} \leq h(p) - h(d(n)) + \frac{\log \log n}{\log n}.$$

This inequality holds if

$$h(d(n)) \leq \frac{\log \log n}{\log n} - O(\frac{\log n}{n}) - \frac{1}{\log n}.$$

It can be seen that for $d(n) = O(\frac{1}{\log n})$, this inequality holds.

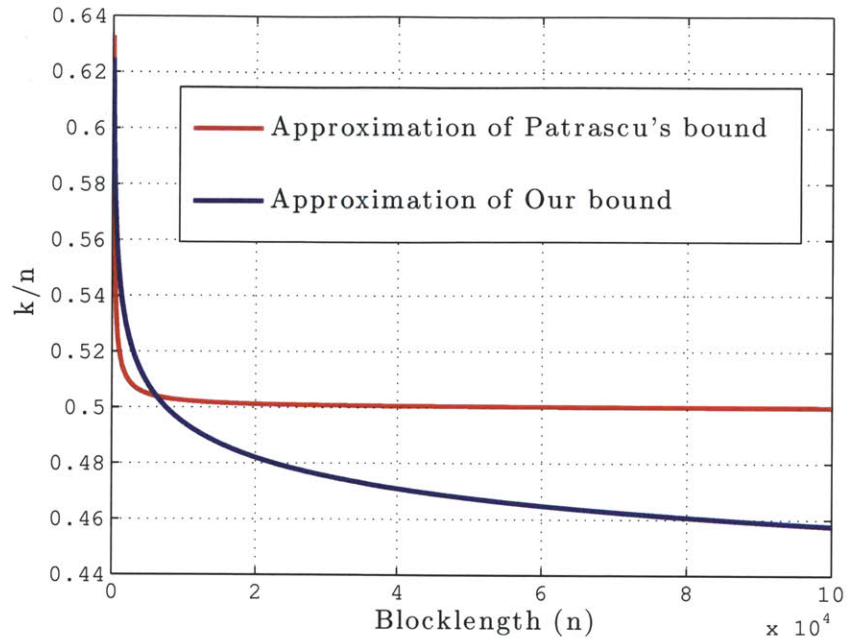

Figure 3-3: Comparison of approximations of upper bounds, $p = .11$, $d = 1/20$

3.2.2 Fixed Number of queries

For a given number of queries, we show that one can achieve any rate above the rate distortion function, using a large enough locality. Consider the following construction. Assume that the source sequence is the i.i.d. product of X with probability measure $\mathbb{P}_X$ ($\mathbb{P}_{X^n} = \mathbb{P}_X^n$). For a given δ, we wish to show that there exists t, such that a LDLSC with locality t achieves the rate $(1+\delta)(R(d))$ with average distortion bounded by d. From Theorem 2, we can get the bound $R_{lsc}(t,d) \leq R(d) + 2\frac{\log t}{t}$ for large enough t. Also let t be large enough such that $2\frac{\log t}{t} \leq \delta R(d)$. Therefore, there exists t such that

$$R_{lsc}(t,d) \leq R(d)(1+\delta).$$

Thus, there exists an encoder and decoder pair for X^t, such that the rate of the code is less than $(1+\delta)R(d)$ and the distortion is bounded by d. Now, consider n pairs of the same encoder-decoder. Concatenate these encoder-decoder pairs to form an encoder-decoder for

X^{nt}. In this way, we obtain a source coding for X^{nt} with distortion

$$\mathbb{E}[\frac{1}{nt}\sum_{i=1}^{nt} d(x_i, \hat{x}_i)] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[d(X^{jt}_{(j-1)t+1}, \hat{X}^{jt}_{(j-1)t+1})] \leq d,$$

and rate

$$R(nt, d) = R(d)(1+\delta).$$

Therefore, there exists a t-local LDLSC with rate $(1+\delta)R(d)$ and average distortion bounded by d for this source.

Proposition 3. *For any source X with probability measure $\mathbb{P}_X$ and any distortion measure, and distortion level, d,*

$$R(d) = \inf\{R \,:\, \exists\, t \text{ and a sequence of } t-\text{ LDLSC with rate } R\}.$$

This proposition states that, in order to achieve the rate $(1+\delta)R(d)$, one need t to be roughly $\frac{1}{\delta R(d)}$.

3.3 LDLSC for Excess Distortion

In rate distortion theory, we usually consider the expected distortion for the best code of block length n and rate less than or equal to R, where $R > 0$ is some fixed number. Instead of this, we shall consider the probability of the event that the distortion exceeds a level $d \geq 0$, if the best code of block length n and rate R is used. We still wish for the decoder to be local. A formal definition is the following.

Definition 10. *A (n, k, d, t, ϵ)-LDLSC for excess distortion is a pair consisting of an encoder $f : X^n \mapsto Y^k$ and a decoder $g : Y^k \mapsto X^n$, where the decoder is t-local. The excess distortion is bounded,*

$$\mathbb{P}[d(X^n, g(f(X^n))) > d] \leq \epsilon.$$

Fix ϵ, d and block-length n. The minimum achievable code size and the finite block length

rate distortion function (excess distortion) are defined by, respectively

$$k^*_{ld}(n,d,t,\epsilon) \triangleq \min\{k \ : \ \exists (n,k,d,t,\epsilon) - \textit{LDLSC for excess distortion}\}, \tag{3.18}$$

$$R_{ld}(n,d,t,\epsilon) \triangleq \frac{k^*_{ld}(n,d,t,\epsilon)}{n}. \tag{3.19}$$

Also, we define

$$R_{ld}(d,t,\epsilon) \triangleq \limsup_{n\to\infty} R_{ld}(n,d,t,\epsilon), \quad R_{ld}(d,t) \triangleq \lim_{\epsilon\to 0} R_{ld}(d,t,\epsilon). \tag{3.20}$$

If we denote the rate of lossy source coding without the local constraint of the decoder by $R(d)$, then it is known that ([4]), for any X with probability measure $\mathbb{P}_X$, the characterization of the rate is as

$$R(d) = \min_{\mathbb{P}_{\hat{X}|X} \,:\, \mathbb{E}[d(X,\hat{X})]\le d} I(X;\hat{X}). \tag{3.21}$$

This means that, for any rate $R \ge \min_{\mathbb{P}_{\hat{X}|X} \,:\, \mathbb{E}[d(X,\hat{X})]\le d} I(X;\hat{X})$, the error goes to zero. The error exponent is given by the following thorem.

Theorem 15 ([14]). *For source with distribution $\mathbb{P}_X$ and a distortion level d, we have: for any ϵ, $\exists\ K_\epsilon$ such that for any $n \ge 0$ there exists encoding-decoding pair f_n and g_n such that*

$$\mathbb{P}[d(g_n(f_n(X^n)), X^n) > d] \le K_\epsilon 2^{-n(F_d(R)-\epsilon)}, \tag{3.22}$$

where,

$$F_d(R) = \min\{D(Q||P) \ : \ R(Q,d) \ge R\}.$$

We use this theorem to design locally decodable codes for excess distortion. Consider a source sequence of length n. Divide the sequence into sequences of length $t(n)$. Therefore, we have $\frac{n}{t(n)}$ blocks of length $t(n)$. Consider the corresponding encoder-decoder pair to each block of length $t(n)$, obtained from Theorem 15. Form an encoder-decoder pair for the whole sequence by concatenating these encoder-decoder pairs. Using the union bound,

we obtain

$$\mathbb{P}[d(g(f(X^n)), X^n) > d] \leq \mathbb{P}[\cup_{j=1}^{\frac{n}{t}}\{d(g(f(X^{jt}_{(j-1)t+1})), X^{jt}_{(j-1)t+1}) > d\}] \leq \frac{n}{t(n)}2^{-t(n)(F_d(R)-\epsilon)}.$$

We need $t(n)$ to be such that

$$\lim_{n\to\infty} \frac{n}{t(n)}2^{-t(n)(F_d(R)-\epsilon)} = 0.$$

Therefore, if

$$t(n) > C \log n,$$

for some C, then the error goes to zero. We have the following result.

Proposition 4. *Let X be a Bern(p) source and $f : X^n \to \{0,1\}^k$ and $g : \{0,1\}^{\to} X^n$ be encoder and $t(n)-$local decoder, respectively. For any $\epsilon > 0$ and $R > R(d)$, there exists a constant C and n_0 such that for $n > n_0$ there exists a $(n, nR, C\log n, d, \epsilon)-$ LDLSC for excess distortion.*

This theorem states that, with a very small number of queries, $O(\log n)$, relative to n, we can achieve any rate above the rate distortion.

3.4 Fixed to Variable Length Local Encoding-Decoding

In this section, we study fixed to variable length source coding, where a sequence of source symbols is mapped to a sequence of bits. We assume that the overall encoder-decoder is local . The definition that follows is motivated by the reference [17]. We call this setting *local encoding-decoding*.

A source code operate as follows. The encoder takes the source sequence $X_1, X_2, ...$ and produces a sequence of bits W, where W is a function of X_1^n. The decoder takes W and produces reproductions $\hat{X}_1, \hat{X}_2, ...$ of $X_1, X_2, ...$ with symbols in a set $\hat{\mathcal{X}}$ called the reproduction alphabet. Note that $\hat{X}_i$ is a function of W. In general, a source code is a system that takes the source symbols $\{X_i\}_{i=1}^n$ and produces $\{\hat{X}_i\}_{i=1}^n$. Therefore, a source code can be characterized by a family of functions $\{g_i\}_{i=1}^n$ called reproduction functions

such that the kth reproduction symbol, $\hat{X}_k$ is

$$\hat{X}_k = g_k(X_1^n), \qquad k = 1, 2, ...,$$

where g_k is a mapping from $\mathcal{X}_1^n$ to $\hat{\mathcal{X}}$.

Definition 11. *A source code $\{g_k\}_{k=1}^n$ is $t-$ local if, for any k, there exists a subset S_k of indices $\{1, 2, ..., n\}$ such that*

$$g_k(x_1^n) = g_k(x'^n_1), \qquad \text{if } x_i = x'_i \text{ for any } i \in S_k,$$

and $|S_k| \leq t$.

Figure 3-4 illustrate the relation between the source and reproduced symbols via a bipartite graph.

Assume a separable distortion function between elements of $\mathcal{X}$ and $\hat{\mathcal{X}}$, $d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^n d(x_i\hat{x}_i)$. When a source code with source code functions $\{g_k\}_{k=1}^n$ is applied to a source X^n, the average distortion is defined as

$$d(\{g_k\}_{k=1}^n) \triangleq \mathbb{E}[d(X^n, \hat{X}^n)]. \tag{3.23}$$

Another measure of the performance of a source code is the rate of code. Let $|W|$ denote the length of W. Therefore, the number of bits for reproducing $\hat{x}_1^n$, for x_1^n as the source

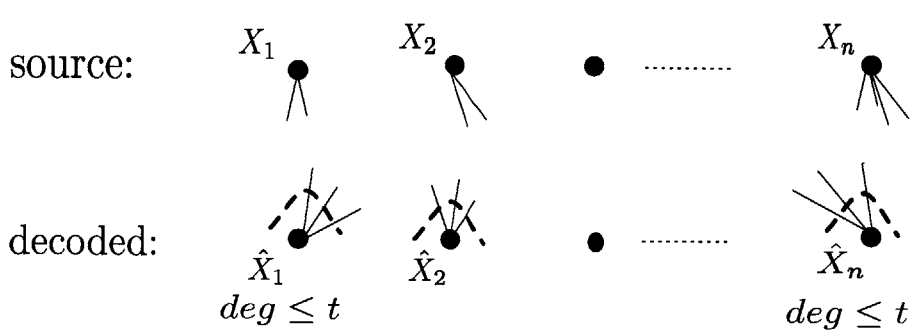

Figure 3-4: Local Encoder-Decoder

sequence is given by

$$L_n(x_1^n) \triangleq |w(x_1^n)|.$$

The average rate of a source code is defined as

$$r(\{g_k\}_{k=1}^n) \triangleq \frac{1}{n}\mathbb{E}[L_n(X^n)]. \tag{3.24}$$

Clearly, there is a tradeoff between average distortion and average rate of a source code. We aim to formulate this tradeoff.

Definition 12. *Assume an i.i.d source with the probability measure $\mathbb{P}_{\mathbb{X}}$. A $(r, D, n, t)_{led}-$ source coding is a pair of encoder-decoder with $t-$local reproduction functions $\{g_i\}_{i=1}^n$ such that the average rate is r, and the average distortion is D. Moreover,*

$$\begin{aligned} r^*_{led}(D, n, t) &\triangleq \min\{r \text{ such that } \exists (r, D, n, t)_{led} - \text{source code}\}. \\ r(D, t)_{led} &\triangleq \limsup_{n\to\infty} r^*_{led}(D, n, t), \end{aligned} \tag{3.25}$$

where the subscript, led, stands for local encoder-decoder.

We may now characterize the average rate for a given source, distortion level (D), and locality (t).

Note that, based on our definition, the average distortion and locality are properties of the reproduction coder functions, $\{g_k\}_{k=1}^n$. Although the average rate is not defined explicitly by the reproduction coder functions, we show the following theorem, that characterizes the average rate in terms of reproduction functions.

Theorem 16. *For a given source, X, with probability measure $\mathbb{P}_X$ and for a given t and distortion level D, we have*

$$r(D, t)_{led} = \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-local \\ d(\{g_k\})\leq D}} \frac{1}{n} H(\hat{X}_1^n) \tag{3.26}$$

Proof: We first prove the converse. For any given $(r, D, n, t)_{led}$-source code, $\hat{X}^n$ is fully represented by a sequence of bits denoted by W. Using, the main result of [1], we

obtain

$$r(n, D, t) \geq \frac{1}{n}\left(H(\hat{X}^n) - \log\left(H(\hat{X}^n) + 1\right) - \log e\right)$$

Since this holds for any source code, taking the infimum of both sides, we obtain

$$r^*_{led}(n, D, t) \geq \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n}\left(H(\hat{X}^n) - \log\left(H(\hat{X}^n) + 1\right) - \log e\right).$$

Taking limit of both sides, we obtain

$$\begin{aligned}
r_{led}(D, t) &\geq \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n}\left(H(\hat{X}^n) - \log\left(H(\hat{X}^n) + 1\right) - \log e\right) \\
&\geq \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n} H(\hat{X}_1^n) - \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n}(\log(H(\hat{X}_1^n) + 1)) \\
&\geq \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n} H(\hat{X}_1^n) - \limsup_{n\to\infty} \frac{\log(n \log|\mathcal{X}| + 1)}{n} \\
&= \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n} H(\hat{X}_1^n). \qquad (3.27)
\end{aligned}$$

To show the achievabilty, consider a $t-$local mapping with the output $\hat{X}^n$. There exists an encoding-decoding of the random variable $\hat{X}^n$ with the average number of bits, not greater than $H(\hat{X}^n)$ ([1]). Denote this encoding and decoding by en and de, respectively. Now, construct the following $(r, D, n, t)_{led}-$source code: the encoder is $en(g(x^n))$ and the decoder the same as de. For this source code, we have $r(D, n, t) \leq \frac{1}{n} H(\hat{X}^n)$. Therefore,

$$r^*_{led}(D, n, t) \leq \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n} H(\hat{X}_1^n)$$

Taking the limit as n goes to ∞ we obtain

$$r_{led}(D, t) \leq \limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\}) \leq D}} \frac{1}{n} H(\hat{X}_1^n) + \epsilon. \quad \square \qquad (3.28)$$

We now analyze the rate of a local source coding by analyzing

$$\limsup_{n\to\infty} \inf_{\substack{\{g_k\}\ t-\text{local} \\ d(\{g_k\})\le D}} \frac{1}{n} H(\hat{X}_1^n).$$

Note that this quantity is only a function of the reproduction coders and does not depend on the encoder that maps the source symbols to bits and the decoder that maps it back to reproduction symbols.

Example 4. *Let X be a uniform random variable on $\mathcal{X} = \{1, 2, 3, 4\}$. Also let $D = \frac{1}{2}$, and $t = 1$, and $d(x, \hat{x}) = 1\{\hat{x} \neq x\}$. We find $r(D, t)$ for this source. First, note that there are only four types of encoder for this source.*

1.

$$W = \begin{cases} 1, & \text{if } X = 1 \\ 1, & \text{if } X = 2, \\ 1, & \text{if } X = 3, \\ 1, & \text{if } X = 4, \end{cases}$$

in this case $H(\hat{X}) = 0$ and $\mathbb{E}[d(\hat{X}, X)] = \frac{3}{4}$.

2.

$$W = \begin{cases} 1, & \text{if } X = 1 \\ 1, & \text{if } X = 2, \\ 1, & \text{if } X = 3, \\ 4, & \text{if } X = 4, \end{cases}$$

in this case $H(\hat{X}) = h(\frac{1}{4})$ and $\mathbb{E}[d(\hat{X}, X)] = \frac{1}{2}$.

3.

$$W = \begin{cases} 1, & \text{if } X = 1 \\ 1, & \text{if } X = 2, \\ 3, & \text{if } X = 3, \\ 4, & \text{if } X = 4, \end{cases}$$

in this case $H(\hat{X}) = H(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ and $\mathbb{E}[d(\hat{X}, X)] = \frac{1}{4}$.

4.

$$W = \begin{cases} 1, & \text{if } X = 1 \\ 2, & \text{if } X = 2, \\ 3, & \text{if } X = 3, \\ 4, & \text{if } X = 4, \end{cases}$$

in this case $H(\hat{X}) = H(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ *and* $\mathbb{E}[d(\hat{X}, X)] = 0$.

Now let α_i *be the fraction of the encoders of the form* i $(1 \leq i \leq 4)$*. Then we have*

$$\mathbb{E}[d(\hat{X}^n, X^n)] = \alpha_1(\frac{3}{4}) + \alpha_2(\frac{2}{4}) + \alpha_3(\frac{1}{4}) + \alpha_4(0),$$

and

$$H(\hat{X}^n) = \alpha_1(0) + \alpha_2(h(\frac{1}{4})) + \alpha_3(\frac{3}{2}\log 2) + \alpha_4(2\log 2).$$

Solving the following linear programming

$$\begin{aligned} \min \quad & \alpha_2(h(\frac{1}{4})) + \alpha_3(\frac{3}{2}\log 2) + \alpha_4(2\log 2) \\ & s.t. \quad \alpha_1(\frac{3}{4}) + \alpha_2(\frac{2}{4}) + \alpha_3(\frac{1}{4}) \leq \frac{1}{2} \\ & \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1, \end{aligned} \tag{3.29}$$

we obtain $\alpha_1 = \frac{2}{3}$ *and* $\alpha_4 = \frac{1}{3}$*, meaning that*

$$r_{led}(D, t) = \frac{2}{3}\log 2.$$

Now consider the following quantity

$$R(D, t) \triangleq \inf_{\substack{\{g_k\}_{k=1}^t \\ d(\{g_k\}) \leq D}} \frac{1}{t} H(\hat{X}_1^t).$$

For any given D*, this quantity gives a corresponding rate. Therefore, we obtain pairs of* (D, R)*. Consider the pair* $(0, 2\log 2)$ *and* $(\frac{3}{4}, 0)$*. With a time sharing between these two points we obtain the point* $\frac{1}{3}(0, 2\log 2) + \frac{2}{3}(\frac{3}{4}, 0)$ *which is the optimal point we obtained before.*

In the following, we generalize this idea to a general setting.

Definition 13. *For a given source, distortion measure, and locality t consider the set of all points (D, R) such that*

$$S(D,R) = \{(D,R) \text{ such that } R = \inf_{\substack{\{g_k\}_{k=1}^t \\ d(\{g_k\}) \leq D}} \frac{1}{t} H(\hat{X}_1^t)\}.$$

Define $\underline{R}(D,t)$ to be the inner convex hull of the points in $S(D,R)$.

The following proposition establishes an upper bound on $r_{led}(D,t)$ in terms of $\underline{R}(D,t)$.

Proposition 5. *For any source, distortion measure, locality t, and distortion level D, we have*

$$r_{led}(D,t) \leq \underline{R}(D,t) \tag{3.30}$$

Proof: Consider a point on $\underline{R}(D,t)$, generated by two points (R_1, D_1) and (R_2, D_2) such that $D = \lambda D_1 + (1-\lambda)D_2$ and $R = \lambda R_1 + (1-\lambda)R_2$ for some $0 \leq \lambda \leq 1$. Also, let g_1 and g_2 be the corresponding reproduction functions of these points. Now consider an encoding-decoding scheme for X^n obtained by dividing X^n into n/t blocks of length t and then applying g_1 to a λ fraction of these blocks and g_2 to the rest of them. By this construction, we get a $t-$local decoder with rate $R = \lambda R_1 + \bar{\lambda} R_2$ and distortion $D = \lambda D_1 + \bar{\lambda} D_2$. Hence, the best rate of $t-$local decoder for this source is below the achieved rate. Therefore,

$$r_{led}(D,t) \leq \underline{R}(D,t).$$

This concludes the proof. □

We conjecture that $\underline{R}(D,t)$ is also a lower bound on $r_{oeld}(D,t)$, meaning that $\underline{R}(D,t) = r_{led}(D,t)$. Note that, this conjecture is closely related to the conjecture stated in Section 3.1.5. The connection is motivated by a class of transformations terms as information preserving transformations, which is defined next. We shall discuss the connection in the next section.

3.5 Information Preserving Transformations

In this section, we introduce a class of information preserving transformations and claim that local transformation are within this class. This class of transformations are quite useful. In particular, this may help us to obtain insight about the conjecture given in Section 3.1.5. Consider two random variables X and Y, where $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. Also, let X^k and Y^k be $k-$fold cartesian products of X and Y, respectively. Let Y^k and X^k be measurable spaces with measure probabilities P_{Y^k} and P_{X^k}, respectively. Let $f : Y^k \mapsto X^k$ be a transformation between these two spaces. We know that, under transformation, entropy is non-increasing i.e.

$$H(Y^k) \geq H(X^k).$$

We are interested in to know the conditions under which the entropy does not decay much. We formulate the problem formally in the following.

Definition 14. *A triple (P_{X^k}, P_{Y^k}, f) is called information preserving if*

$$H(Y^k) \leq H(X^k) + o(k). \tag{3.31}$$

We now study the information preserving transformations. Mainly, we propose a class of local transformations and conjecture that, with some assumptions on the measure P_{X^k}, they are information preserving.

Definition 15. *A transformation $f : Y^k \mapsto X^k$ is $t-$local if for any $1 \leq i \leq k$, X_i is a function of only t of the components of Y^k.*

First, we illustrate by an example the reasoning for the assumptions we shall have.

Example 5. *In this example we illustrate that only being local is not sufficient for a transformation to be information preserving. We also need some assumptions on the resulting probability distribution on X^k. Let $X_i = Y_1$ for all $1 \leq i \leq k$, then $H(X^k) = H(Y_1) << H(Y^k)$.*

Therefore, we assume that the resulting distribution of X^k is very close to i.i.d in the following sense.

Definition 16. *A probability measure, P_{X^n}, on X^n is called ϵ-i.i.d. if $||P_{X^n} - I_{X^n}|| \leq \epsilon$, where $||.||$ denotes the total variation distance between two measures and I_{X^n} is some i.i.d measure on X^n.*

Note again that these assumptions are not sufficient for a local mapping to be information preserving. Next example shows another necessary aassumption.

Example 6. *Let $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3\}$. Assume X^k is i.i.d with distribution $P_X = (1/2, 1/2, 0, 0)$. Also, let Y_i to be defined as the following*

$$Y_i = \begin{cases} 0, & \text{if } X_{2i-1} = X_{2i} = 0 \\ 1, & \text{if } X_{2i-1} = 1, X_{2i} = 0 \\ 2, & \text{if } X_{2i-1} = 0, X_{2i} = 1 \\ 3, & \text{if } X_{2i-1} = X_{2i} = 1. \end{cases}$$

$Y^{\frac{k}{2}}$ is then distributed i.i.d according to $P_Y = (1/4, 1/4, 1/4, 1/4)$ (assume k is an even integer number). Let Y^is for $\frac{k}{2} + 1 \leq i \leq k$ be distributed i.i.d with the same probability measure. X^k is obtained by a local mapping of $Y^{\frac{k}{2}}$, where $t = 1$. On the other hand, we have $H(X^k) = k \log 2$ and $H(Y^k) = k \log 4$, therefore, we have $H(X^k) << H(Y^k)$.

The issue of this example is that, P_X is degenerate. In general, even if the distributions are not degenerate, the sets $\mathcal{X}$ and $\mathcal{Y}$ play an important role in this definition of information preserving transformation. For instance, if we assume that $|\mathcal{Y}|$ is much larger than $|\mathcal{X}|$, then we can always construct an example similar to Example 6. Consider the following example.

Example 7. *Let X be some non-degenerate random variable on $\mathcal{X}$. Let $\mathcal{Y} = \mathcal{X} \times \mathcal{X} \times \mathcal{X}$. Now, we have a local transformation from Y^k to X^k, while $H(X^k) = kH(X)$ and $H(Y^k) = 3kH(X) >> H(X^k)$.*

Therefore, we must assume that $\mathcal{X}$ and $\mathcal{Y}$ have the same cardinality and both random variables X and Y are non-degenerate. We have the following conjecture on information preserving mappings:

Let $|\mathcal{X}| = |\mathcal{Y}|$. For small enough ϵ, if P_{X^k} is ϵ-i.i.d. and f is a $t-$ local map, then

(P_{X^k}, P_{Y^k}, f) is an information preserving mapping.

Assuming this conjecture is true, one can work out to prove the conjecture given in Section 3.1.5 on the rate of LDSC with fixed number of queries. Moreover, the conjecture of the previous section is the the analogy of this conjecture for lossy setting.

3.6 Storage on Memories with Access Cost

In this section, we study a problem of storing source sequence on two memories with different types as defined in the following. We have two different types of memories. One is the original memory where the bits are stored, denoted by M_o, and the other one is a cache memory, denoted by M_c. We Also know that the cost of querying one bit from M_o is higher than the cost of querying from M_c (cache memory is faster). Moreover, the cost of occupying M_o is less than the cost of occupying M_c (cache memory is rare). The goal is to find the trade-off between cost of occupying the memories and cost of querying the memories in the reconstruction of the source.

Let $f : X^n \mapsto \{0,1\}^k$ be the encoder. The decoder is a set of n functions $g_i : \{0,1\}^k \times X^n \mapsto \{0,1\}$, where g_i denotes the ith coordinate of the overall decoding function, g, for $1 \leq i \leq n$. g_i decodes $\hat{X}_i$, the reproduction of X_i. Assume the cost of querying the original memory is C_o and the cost of querying cache memory is C_c. We wish to have a bounded recovery error $\mathbb{P}[X^n \neq \hat{X}^n] \leq \epsilon$. This setting is called a (n, k, C, ϵ)-cost code, where the average cost of querying is bounded by a given budget, C. In other words, if t_i and $\hat{t}_i$ denote the number of queries g_i asks from $\{0,1\}^k$ and X^n, respectively. The cost of g_i is then $C_o\hat{t}_i + C_c t_i$ and the cost of the whole decoder is $\sum_{i=1}^n C_o\hat{t}_i + C_c t_i$. Note that if we wanted to ask some queries from X^n to recover X_i, the we only need one query because X^n is stored on M_o. Therefore, we either have $\hat{t}_i = 1$ and $t_i = 0$ or $\hat{t}_i = 0$ and $t_i \geq 1$. We seek to reduce the cache usage under the assumption that in the storage system, cost of using cache memory is much higher than non-cache memory. The following is the formal definition of the code rate for this problem.

Definition 17. *Let* $k^*_{cost}(n, \epsilon, C) \triangleq \min\{k$ *such that* $\exists (n, k, C, \epsilon)$*-cost code*$\}$,

and

$$R_{cost}(n, \epsilon, C) = \frac{k^*_{cost}(n, \epsilon, C)}{n}.$$

Also, define the rate as

$$R_{cost}(C) = \lim_{\epsilon \to 0} R_{cost}(\epsilon, C),$$

where

$$R_{cost}(\epsilon, C) = \limsup_{n \to \infty} R_{cost}(n, \epsilon, C).$$

First, note that in the case where the cost of reading from original memory is C_o and the cost of reading from cache memory is $C_c = 0$, the rate is $R_{cost}(C) = (1 - \frac{C}{C_o})h(p)$. We now consider the case where the cost of reading from cache memory is not zero. In this case we can normalize all the costs and assume the cost of reading from cache memory is 1 and the cost of reading from original memory is C_o.

Proposition 6. *For any* $0 < \lambda < 1$, *we have the following relationship between a cost code and ALDSC:*

$$k^*_{cost}(n, \epsilon, \bar{\lambda} C_o + \lambda t) \leq k^*_{ald}(\lambda n, \epsilon, t).$$

Proof: Let $(\lambda n, k, \epsilon, t)$ be an ALDSC with encoder f and decoder g. Form an encoding-decoding for storage problem as follows. Use the same encoding function. Assume g'_i is the same as g_i for X_i, where $1 \leq i \leq \lambda n$ and is identical mapping form X^n for $\lambda n \leq i \leq n$. We have

$$\frac{1}{n}\bar{\lambda} n C_o + \frac{1}{n}\sum_{i=1}^{\lambda n} t_i \leq \bar{\lambda} C_o + \lambda t,$$

also the error probability is still bounded by ϵ. Thus, we find a $(n, k, \bar{\lambda} C_o + \lambda t, \epsilon)-$ cost code. Therefore,

$$\{\, k \mid \exists(\lambda n, k, \epsilon, t)\text{-ALDSC}\} \subseteq \{\, k \mid \exists(n, k, \bar{\lambda} C_o + \lambda t, \epsilon)\text{-cost code}\} \tag{3.32}$$
$$\Rightarrow k^*_{cost}(n, \epsilon, \bar{\lambda} C_o + \lambda t) \leq k^*_{ald}(\lambda n, \epsilon, t)$$

Corollary 7. *For any* $0 \leq \lambda \leq 1$, *we have* $R_{cost}(\bar{\lambda} C_o + t\lambda) \leq \lambda R_{ald}(t)$.

Proof: Using Proposition 6, we obtain

$$\frac{k^*_{cost}(n, \epsilon, \bar{\lambda}C_o + t\lambda)}{n} \leq \lambda \frac{k^*_{ald}(\lambda n, \epsilon, t)}{\lambda n}$$
$$R_{cost}(\epsilon, \bar{\lambda}C_o + t\lambda) \leq \lambda R_{ald}(\epsilon, t)$$
$$R_{cost}(\bar{\lambda}C_o + t\lambda) \leq \lambda R_{ald}(t).$$

Proposition 7. *There exists* $0 \leq \lambda \leq 1$ *for which*

$$k^*_{ald}\left(\bar{\lambda}n, \epsilon, \frac{C - \lambda C_o}{\bar{\lambda}}\right) \leq k^*_{cost}(n, \epsilon, C)$$

Proof: Consider a (n, k, ϵ, C)-cost code. Assume λ fraction of bits are recovered from memory and the rest of them are recovered from cache memory. Without loss of generality, assume the first λ fraction are recovered directly from the memory and the rest of the bits are recovered by t_i queries ($(1-\lambda)n \leq i \leq n$). Therefore, $C = \lambda C_o + \frac{(1-\lambda)}{(1-\lambda)n} \sum_{i=1}^{(1-\lambda)n} t_i$. Thus, we can extract a $((1-\lambda)n, k, \epsilon, \frac{C-\lambda C_o}{(1-\lambda)})$ ALDSC. An argument similar to the proof of Proposition 6 completes the proof.

Corollary 8. *There exists* $0 \leq \lambda \leq 1$ *for which*

$$\bar{\lambda} R_{ald}(\frac{C - \lambda C_o}{(1-\lambda)}) \leq R_{cost}(C).$$

Moreover, because of the cost constraint we have

$$\lambda C_o + (1-\lambda)C_c \leq C.$$

Proof: The proof follows directly form Proposition 7.

Theorem 17. *Using linear encoder, If* $C_c > 0$, *then* $R_{cost}(C) = \frac{C-C_o}{C_c-C_o}$.

Proof: Using Corollary 8 and the fact that, the rate of ALDSC with using linear encoder is 1, we have $R_{cost}(C) \geq (1-\lambda)$ such that $(1-\lambda)R_{ald}(\frac{C-\lambda C_o}{(1-\lambda)}) \leq R_{cost}(C)$. Solving this minimization problem to find the smallest $(1-\lambda)$, we obtain $R_{cost}(C) = \frac{C-C_o}{C_c-C_o}$ which concludes the proof. □

Assume a storage problem with two types of memories when one of the memories is a cache memory and the other one is a regular memory. The result of this section shows that, the optimal policy to reduce the cost of recovery is to store as much bits as the budget, C, allows on the cache memory without encoding, and then store the rest of bits in the non-cache memory. Therefore, designing any encoder-decoder function for this problem does not give a better result than the naive way of storing the bits without coding.

Chapter 4

Conclusion and future work

We have formulated and studied the problem of source coding with some constraints on the encoder-decoder. Based on the results in the literature, when the encoder is local, there exist codes that achieve fundamental entropy rate with and without linear encoders. Also, the results in the literature characterize the tradeoff between locality of encoder and the rate of a lossy source coding. The focus of this work is on the source coding with local decoder. The following summarizes the main results we showed in this work:

- Almost lossless source coding:
  - Constant locality: It is shown that the rate of source coding with linear encoder/decoder is one, meaning that no compression is possible. Also for small locality ($t = 2$), the rate of any encoder-decoder is one. One application of these results is given in the context of data storage management.
  - Scaling locality: Any given rate above the Shannon fundamental entropy rate is achievable with logarithmic locality in the block-length.
- Lossy source coding:
  - Constant locality: Any given rate above the Shannon fundamental rate distortion is achievable with a proper constant locality. This locality is proportional to the inverse of the distance between the given rate and rate distortion.

- Scaling locality: Any given rate above the Shannon fundamental rate distortion is achievable with any scaling locality ($\lim_{n\to\infty} t(n) = \infty$) and the rate of convergence is upper bounded as in Theorem 13. Moreover, for lossy source coding with excess distortion, logarithmic locality suffices to achieve any rate above the rate distortion.
- We formulated fixed to variable local source coding and characterized the rate. We also established an upper bound on the rate.

There are connections between locally decodable codes and finitary codes [10] which requires more exploration. The locally decodable source coding applied to non i.i.d sources such as Markov sources is another future topic to be studied.
There are several interesting unsolved problems about local codes.

- We proved converse bound on the rate of LDSC with fixed number of queries. We established this converse for the following cases: linear encoder, linear decoder, and general encoder-decoder with locality $t = 2$. We conjectured that for any general encoder-decoder and any t, the converse bound holds .
- We introduced the class of information preserving mappings. We then, conjectured that local mappings are in this class. Proving that this conjecture holds, shows that the previous conjecture also holds.
- We introduced fixed to variable local encoding-decoding. We formulated the rate of this source code, which is given in Theorem 16. We conjectured that, asymptotically the best $t-$local source code is given by combining two source codes which are designed for source sequences of length t. This conjecture establishes a converse bound on the rate of fixed to variable local coding. The achievability bound is given in Proposition 5. Note that these two bounds are equal, implying that we can characterize the rate of the code.

Bibliography

[1] Noga Alon and Alon Orlitsky. A lower bound on the expected length of one-to-one codes. *Information Theory, IEEE Transactions on*, 40(5):1670–1672, 1994.

[2] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, 1970.

[3] V. Chandar, D. Shah, and G.W. Wornell. A locally encodable and decodable compressed data structure. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 613–619. IEEE, 2009.

[4] Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley-interscience, 2012.

[5] Imre Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *Information Theory, IEEE Transactions on*, 28(4):585–592, 1982.

[6] Imre Csiszar and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[7] AG Dimakis, MJ Wainwright, and K Ramchandran. Lower bounds on the rate-distortion function of ldgm codes. In *Information Theory Workshop, 2007. ITW'07. IEEE*, pages 650–655. IEEE, 2007.

[8] G. Jacobson. Space-efficient static trees and graphs. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 549–554. IEEE, 1989.

[9] Y. Kaspi and N. Merhav. Zero-delay and causal single-user and multi-user lossy source coding with decoder side information. *CoRR*, abs/1301.0079, 2013.

[10] M Keane and M Smorodinsky. A class of finitary codes. *Israel Journal of Mathematics*, 26(3-4):352–371, 1977.

[11] V. Kostina and S. Verdú. Fixed-length lossy compression in the finite blocklength regime. *Information Theory, IEEE Transactions on*, 58(6):3309–3338, 2012.

[12] S. Kudekar and R. Urbanke. Lower bounds on the rate-distortion function of individual ldgm codes. In *Turbo Codes and Related Topics, 2008 5th International Symposium on*, pages 379–384. IEEE, 2008.

[13] D.J.C. MacKay. Good error-correcting codes based on very sparse matrices. *Information Theory, IEEE Transactions on*, 45(2):399–431, 1999.

[14] K. Marton. Error exponent for source coding with a fidelity criterion. *Information Theory, IEEE Transactions on*, 20(2):197–199, 1974.

[15] A. Mazumdar, G. W Wornell, and V. Chandar. Update efficient codes for error correction. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1558–1562. IEEE, 2012.

[16] A. Montanari and E. Mossel. Smooth compression, gallager bound and nonlinear sparse-graph codes. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 2474–2478. IEEE, 2008.

[17] D. Neuhoff and R. Gilbert. Causal source codes. *Information Theory, IEEE Transactions on*, 28(5):701–713, 1982.

[18] R. Pagh. Low redundancy in static dictionaries with constant query time. *SIAM Journal on Computing*, 31(2):353–363, 2001.

[19] D.S. Papailiopoulos and A.G. Dimakis. Locally repairable codes. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2771–2775. IEEE, 2012.

[20] M. Patrascu. Succincter. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 305–313. IEEE, 2008.

[21] Y. Polyanskiy, H.V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. volume 56, pages 2307–2359. IEEE, 2010.

[22] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 4(142-163), 1959.

[23] C. E. Shannon, W. Weaver, R. E. Blahut, and B. Hajek. *The mathematical theory of communication*, volume 117. University of Illinois press Urbana, 1949.

[24] W. Szpankowski and S. Verdú. Minimum expected length of fixed-to-variable lossless compression without prefix constraints. *Information Theory, IEEE Transactions on*, 57(7):4017–4025, 2011.

[25] Lav R Varshney, J. Kusuma, and V. K. Goyal. Malleable coding: Compressed palimpsests. *arXiv preprint arXiv:0806.4722*, 2008.

[26] S. Yekhanin. New locally decodable codes and private information retrieval schemes. In *Electronic Colloquium on Computational Complexity, vol. TR06*, page 127, 2006.

[27] Z. Zhang, E.H. Yang, and V.K. Wei. The redundancy of source coding with a fidelity criterion. 1. known statistics. *Information Theory, IEEE Transactions on*, 43(1):71–91, 1997.

[28] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *Information Theory, IEEE Transactions on*, 24(5):530–536, 1978.