論文 / 著書情報
Article / Book Information

題目(和文)	個別系列の複雑量とユニバーサル情報源符号化への応用
Title(English)	Complexities of individual sequences and their applications to universal source coding
著者(和文)	葛岡成晃
Author(English)	Shigeaki Kuzuoka
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第6828号, 授与年月日:2007年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:甲第6828号, Conferred date:2007/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Complexities of Individual Sequences and Their Applications to Universal Source Coding

Shigeaki KUZUOKA

Supervisor: Prof. Tomohiko UYEMATSU

Department of Communications
and Integrated Systems,
Tokyo Institute of Technology

2007

Acknowledgements

First of all, I wish to express my sincere gratitude and special thanks to my supervisor, Professor Tomohiko Uyematsu for his constant guidance and close supervision during all the phases of this work. Without his guidance and valuable advice, I could not accomplish my research.

I would also wish to express my gratitude to Prof. Ryutaroh Matsumoto for valuable advice and support.

It is my pleasure to deeply thank Prof. Te Sun Han for fruitful comments and suggestions. His book in collaboration with Prof. Kingo Kobayashi [8] gave great influence on me when I was an undergraduate student.

Constructive comments and suggestions given at conferences and seminars have significantly improved the presentation of my result. Especially, I would like to thank Prof. Hirosuke Yamamoto, Prof. Hiroshi Nagaoka, Prof. Tsutomu Kawabata, Prof. Yasutada Oohama, Prof. Ken-ichi Iwata, Prof. Hiroki Koga, Dr. Jun Muramatsu, Dr. Mitsuharu Arimura, and Dr. Mitsugu Iwamoto for valuable comments. I would also like to thank anonymous reviewers of [16] and [18] for their useful comments.

I would also like to thank Prof. Yoshinori Sakai, Prof. Koichi Sakaniwa, and Prof. Isao Yamada for giving valuable advice.

My deep thanks are also addressed to all my colleagues at Uyematsu-laboratory for their support and helpful comments during regular seminars. I would also like to thank our group secretary Ms. Junko Goto for her care and kindness.

Finally, I would like to deeply thank my parents for their support and encouragement.

Contents

Chapter 1

Introduction

The development of computer communication networks is bearing massive transfer of various kinds of data over networks, and many data processing applications require storage of a large volume of data. To reduce the consumption of expensive resources, such as transmission bandwidth or disk space, various kinds of techniques of data compression have been proposed. Further, the practical requirement of data compression has yielded a large volume of research on source coding problems.

Claude Shannon initiated the theory of lossless source coding problem (along with other subjects in the information theory) in the ground-breaking paper [23]. Shannon formulated the concept of entropy, and gave it operational significance by proving the source coding theorem. His theorem states that the entropy rate of a stationary ergodic Markov source with finite alphabet is the optimum rate at which the source can be encoded losslessly. Later, extensions and generalizations of Shannon's theorem were obtained. Today, it is known that the source coding theorem holds for extremely generalized sources which do not satisfy stationarity nor compatibility [7].

While a lossless code encodes the data so that the original data can be reconstructed, lower compression rate can be achieved if some distortion between the original data and reconstructed data is admitted. In 1948, Shannon [23] defined the concept of rate-distortion function and then sketched the proof of a lossy source coding theorem. Later, Shannon [24] formulated the rate-distortion theory, where the optimal trade-off between the distortion and the coding rate is investigated. Shannon's lossy coding theorem states that the rate-distortion function for a discrete memoryless source gives the optimum compression rate at which the source can be encoded so that the resulting distortion in the recovery of the source output is less than the given distortion level. A large number of papers has been devoted to the rate-distortion theory [13]. Three kinds of lossy coding problems were formulated, namely, fixed-rate, fixed-distortion, and fixed-slope lossy coding problems. In the fixed-rate (resp. fixed-distortion) coding problem, one minimizes the distortion (resp. coding rate) subject to a constraint on the coding rate (resp. distortion). In addition to the fixed-rate and fixed-distortion coding, the fixed-slope lossy coding problem of stationary ergodic sources was first proposed by Yang, Zhang, and Berger [33]. In the fixed-slope coding problem, neither the rate nor the distortion needs to be fixed, and instead, one minimizes the cost defined by the rate and the distortion.

It has been recognized that the theories of lossless and lossy source coding provide a theoretical basis for many practically important data compression problems. On the other hand, in order to design an optimal code, in the source coding theorems, it is necessary to know the stochastic properties of the source. In other words, the statistical properties of the source should be fully investigated in advance of constructing the code. However, in a real situation, estimating the statistical properties of the source may be either impossible or unreliable. Moreover the stochastic properties of source may vary with time in an unpredictable way.

One way to treat the uncertainty on stochastic properties of a source is to consider a class of sources rather than a particular source. A code is said to be universal for a class of sources if, for any source in the class, there is no asymptotic loss of performance when we use the code instead of an optimal code for the source. For some classes of sources, various kinds of universal lossless code have been proposed [37, 4, 10, 30]. Moreover, even in lossy coding problems, there are various kinds of universal codes for some classes of sources [38, 10, 21].

Another way to treat the uncertainty on stochastic properties of a source is to regard the data as an individual sequence rather than an output from a stochastic source. Nonprobablistic measure of the compressibility of individual data can be defined as the smallest compression rate achievable by a given class of codes. Kolmogorov [15], Chaitin [1], and Solomonoff [26] considered the class of codes that output a binary program which causes the original input data when run on a Turing machine. The resulting measure is called Kolmogorov complexity. Ziv and Lempel [36] considered the class of codes which are modeled by finite-state machines. They revealed that, in the theory of coding problem of individual sequences, the compressibility of sequences plays a role analogous to that of entropy in Shannon's source coding theorem. In addition, their study yielded a universal lossless code which is as known as Lempel-Ziv 78 (LZ78) code. On the other hand, the framework of lossless coding of individual sequences was also extended to the lossy coding problem of individual sequences. Ziv [34] considered the lossy coding problem of individual sequences, and proposed a fixed-rate universal lossy code. Yang and Kieffer [31] extended Ziv's result, and investigated the fixed-rate and fixed-distortion lossy coding problems of individual sequences. They also showed that LZ78 code can be applied to universal lossy coding.

In this thesis, we adopt the framework of coding problems of individual sequences. The aim of this thesis is to reveal relations among various kinds of complexities of sequences and investigate coding problems of individual sequences in a unified way.

First, we consider the lossless coding problem of individual sequences over countably infinite alphabet. We investigate relations among two kinds of empirical entropies, the self-entropy rate, and the Lempel-Ziv complexity of individual sequences over countably infinite alphabet. Using these complexities, we characterize the compressibility of individual sequences via the class of blockwise encoders. Moreover, as an application of this result, we show that a variation of LZ78 code can attain a better compression rate than any blockwise encoder, even if the alphabet is countably infinite.

Next, we investigate the fixed-rate and fixed-distortion lossy coding problems of individual sequences. We deal with more general distortion measures than the additive distortion measure considered in [31]. For each problem, we characterize the optimal performance theoretically attainable via blockwise lossy encoders by using empirical distributions of individual sequences. In addition, we propose the fixed-rate and fixed-distortion universal lossy encoders based on the complexity of sequences. This result reveals that not only LZ78 code, but various universal lossless codes such as grammar based codes can also be applied to universal lossy coding.

Finally, we investigate the fixed-slope lossy coding problem of individual sequences. We show that the optimal cost attainable via blockwise lossy encoders can be characterized by using two kinds of empirical distributions of sequences. Moreover, we apply this result to the fixed-slope lossy coding problem of nonstationary sources.

Our results show that optimal performances attained in lossless and lossy coding problems are characterized by complexities of individual sequences. Coding theorems clarify that the empirical entropy (resp. distortion-rate function, rate-distortion function, cost function) of the sequence gives the optimal compression rate (resp. distortion, coding rate, cost) attainable by blockwise encoders. This fact means that when we consider coding problems of individual sequences, these complexities play important roles analogous to that of quantities defined for probabilistic sources. Though we cannot obtain exact values of complexities in general, these complexities have an advantage that they are defined for each individual sequence independently of probabilistic sources. Especially, some complexities are defined by using the overlapping empirical distribution of the sequence. By combining the ergodic theorem, our results on the coding problems of individual sequences can be applied to the universal coding problems of stationary ergodic sources. It can be shown that for almost all sequence emitted from a stationary ergodic source, the empirical entropy (resp. distortion-rate function, rate-distortion function, cost function) of the sequence is equal to the entropy rate (resp. distortion-rate function, rate-distortion function, cost function) of the source.

This thesis is organized as follows. In Chapter 2, we define the notation and terminology used in this thesis, and then, investigate the lossless coding problem of individual sequences. In Chapter 3 and Chapter 4, we investigate the fixed-rate and fixed-distortion lossy coding problem of individual sequences respectively. In Chapter 5, we investigate the fixed-slope lossy coding problem of individual sequences and nonstationary sources. In Chapter 6, we summarize the result of this thesis.

Chapter 2

Lossless Coding Problem

2.1 Introduction

Data compression has important applications in the areas of data transmission and data storage. The development of computer communication networks is bearing massive transfer of data over networks, and many data processing applications require storage of a large volume of data. It is inappropriate to transmit or store a large amount of data in its original form. If we analyze and remove the redundancy contained in the original data, the amount of data to be transmitted or stored becomes small, and thus, the cost for transmission or storage can be reduced.

In the lossless coding problem of individual sequences, we consider the following mathematical model of a data compression system. When we wish to transmit or store a given data sequence, we use an encoder to convert the sequence into the binary form called the codeword. From the codeword, the decoder reproduces the original data. Since it is necessary to guarantee that the original data sequence can be reproduced from the codeword, we cannot assign short codewords to all sequences. In this chapter, we investigate how well we can losslessly compress a given individual sequence.

Historically, the lossless coding problem of individual sequences was formulated by Ziv and Lempel [19, 35, 36, 39]. Lempel and Ziv [19] investigated the *complexity* of finite sequences, and then later, they applied their result to a universal algorithm for sequential data compression [35]. Moreover, Ziv and Lempel [36] revealed the compressibility of individual sequences by the class of finite-state information-lossless encoders. Ziv [39] also established the coding theorems for fixed-rate lossless coding of individual sequences. In

Section 2.3, we review the Ziv and Lempel's result [36] on the lossless coding of individual sequences via finite-state encoders.

While Ziv and Lempel considered the lossless coding problem of sequences over finite alphabet, only little is known about the lossless coding problem of sequences over countably infinite alphabet. In Section 2.4, we consider the lossless coding problem of individual sequences over countably infinite alphabet, and characterize the compressibility of an individual sequence via blockwise lossless encoders. Our main theorem shows that, for a given sequence, the optimal performance attainable by blockwise coding is equal to the self-entropy rate and the empirical entropies. Moreover, it is also shown that the Lempel-Ziv complexity, which is defined by the minimum number of phrases in arbitrary distinct parsing of the sequence, gives a lower bound of the compression rate. As an application of this result, in Section 2.6, we reveal that a variation of LZ78 code can attain a better compression rate than any block encoder. Further, we clarify a relation between our result and Kieffer's result [10] on the universal coding of the stationary ergodic sources.

2.2 Notation and Terminology

Throughout this thesis, we use the following notation and terminology.

Let A be a set. We denote by $|A|$ the cardinality of A. We denote by A^n the set of all strings formed from n symbols in A, and by A^∞ the set of all infinite sequences over A. The set $A^* \triangleq \bigcup_{n=0}^{\infty} A^n$ denotes the set of all strings of finite length. For each infinite sequence $a = a_1 a_2 \cdots \in A^\infty$ ($a_i \in A$ for each $i = 1, 2, \cdots$) and two positive integers n and m ($n \leq m$), a_n^m denotes the subsequence $a_n a_{n+1} \cdots a_m$ of a. For two strings $u \in A^*$ and $v \in A^*$, uv denotes the concatenation of strings.

Let $\mathcal{N}$ be the set of positive integers $\{1, 2, 3, \cdots\}$ and $\mathcal{B} \triangleq \{0, 1\}$ be the binary set. The function $\ell : \mathcal{B}^* \to \mathcal{N}$ denotes the length of the binary string, that is, $\ell(b_1^n) = n$ for each binary sequence $b_1^n = b_1 b_2 \cdots b_n \in \mathcal{B}^n$.

For any nonnegative number $\tau \geq 0$, $\lceil \tau \rceil$ denotes the smallest integer greater than or equal to τ, and $\lfloor \tau \rfloor$ denotes the largest integer smaller than or equal to τ.

Throughout this thesis, we take all logarithms to the base 2. We adopt the convention that $0 \log 0 = 0$, which is justified by continuity since $\tau \log \tau \to 0$ as $\tau \to 0$.

2.3 Lossless Coding of Sequences over Finite Alphabet

In this section, we review the Ziv and Lempel's result on the lossless coding problem of individual sequences over finite alphabet. Throughout this section, we assume that $\mathcal{X}$ is a set of finite symbols. $\mathcal{X}$ is called the *source alphabet*. Ziv and Lempel [36] investigated the compressibility of individual sequences with respect to codes that can be modeled by finite-state automata. In their model, an encoder ψ is defined by a quintuple $(\mathcal{S}, \mathcal{X}, B, g, f)$ where $\mathcal{S}$ is a finite set of states, $B \subset \mathcal{B}^*$ is a finite set of binary words including null-word, $g : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$ is the next-state function, and $f : \mathcal{S} \times \mathcal{X} \to \mathcal{B}^*$ is the output function. When an infinite sequence $x = x_1 x_2 \cdots$ $(x_i \in \mathcal{X})$ is fed into the encoder $\psi = (\mathcal{S}, \mathcal{X}, B, g, f)$, the encoder ψ emits an infinite sequence $u = u_1 u_2 \cdots$ $(u_i \in B)$ while going through an infinite sequence of states $s = s_1 s_2 \cdots$ $(s_i \in \mathcal{S})$, according to

$$\begin{aligned} u_i &= f(s_i, x_i) \\ s_{i+1} &= g(s_i, x_i) \quad i \in \mathcal{N}, \end{aligned}$$

where s_i is the state of ψ when it processes the input sequence x_i. An encoder ψ is said to be *information lossless* (IL) if for all $s_1 \in \mathcal{S}$ and all $x_1^n \in \mathcal{X}^n$ $(n \in \mathcal{N})$, the triple $\{s_1, u_1^n, s_{n+1}\}$ uniquely determines x_1^n. In the following discussion, we assume that the initial state s_1 is a prescribed fixed member of the state-set $\mathcal{S}$.

Given an encoder $\psi = (\mathcal{S}, \mathcal{X}, B, g, f)$ and an input string $x_1^n \in \mathcal{X}^n$, the *compression rate* for x_1^n with respect to ψ is

$$\frac{\ell(\psi(x_1^n))}{n},$$

where $\psi(x_1^n) \equiv f(s_1, x_1) f(s_2, x_2) \cdots f(s_n, x_n)$ is the output binary string when x_1^n is encoded by ψ. Then, for each individual sequence $x \in \mathcal{X}^\infty$, the *finite-state compressibility* of x is defined as the optimal performance theoretically attainable (OPTA) in lossless compression via finite-state IL encoders.

Definition 2.1 (finite-state compressibility [36]) For a given sequence $x \in \mathcal{X}^\infty$, let

$$\rho(x) \triangleq \lim_{S \to \infty} \limsup_{n \to \infty} \min_{\psi \in \mathcal{E}(S)} \frac{\ell(\psi(x_1^n))}{n},$$

where $\min_{\psi\in\mathcal{E}(S)}$ is taken over all finite-state IL encoders with $|\mathcal{S}| \leq S$.

The real-valued function $x \mapsto \rho(x)$ is referred as the *OPTA function* in lossless coding. To characterize the finite-state compressibility $\rho(x)$, we introduce two kinds of quantities which indicate *complexities* of a sequence.

Definition 2.2 (Overlapping empirical entropy [36]) The *overlapping empirical distribution* $p_k(\cdot|x_1^n)$ of x is defined as

$$p_k(a_1^k|x_1^n) \triangleq \frac{|\{i : 0 \leq i \leq n-k, x_{i+1}^{i+k} = a_1^k\}|}{n-k+1} \quad \forall a_1^k \in \mathcal{X}^k. \tag{2.1}$$

Let

$$\hat{H}_k(x_1^n) \triangleq -\frac{1}{k} \sum_{a_1^k \in \mathcal{X}^k} p_k(a_1^k|x_1^n) \log p_k(a_1^k|x_1^n),$$

and

$$\hat{H}_k(x) \triangleq \limsup_{n\to\infty} \hat{H}_k(x_1^n).$$

Then, the *overlapping empirical entropy* $\hat{H}(x)$ of $x \in \mathcal{X}^\infty$ is defined as

$$\hat{H}(x) \triangleq \inf_k \hat{H}_k(x).$$

Definition 2.3 (Lempel-Ziv complexity [36]) For all $x \in \mathcal{X}^\infty$, let

$$C(x) \triangleq \limsup_{n\to\infty} \frac{\log n}{n} c(x_1^n),$$

where $c(x_1^n)$ denotes the maximum number of phrases in arbitrary distinct parsing of x_1^n.

Ziv and Lempel showed that the finite-state compressibility $\rho(x)$ is equal to $\tilde{H}(x)$ and is lower bounded by $C(x)$.

Theorem 2.1 ([36]) If $\mathcal{X}$ is finite then for any individual sequence $x \in \mathcal{X}^\infty$,

$$C(x) \leq \tilde{H}(x) = \rho(x).$$

2.4 Lossless Coding of Sequences over Countably Infinite Alphabet

In the remaining part of this chapter, we assume that the source alphabet $\mathcal{X}$ is countably infinite set. It is clear that any finite-state encoder does not have enough memory to deal with infinite alphabet. Let $\psi = (\mathcal{S}, \mathcal{X}, B, g, f)$ be a finite-state encoder. Since $\mathcal{S}$ and B are finite, there exists $a \in \mathcal{X}$ and $a' \in \mathcal{X}$ such that $f(a,s) = f(a',s)$ and $g(a,s) = g(a',s)$ for any $s \in \mathcal{S}$, and thus, any triple $\{s, f(a,s), g(a,s)\}$ cannot determine a uniquely. To formulate the problem properly, we adopt blockwise lossless encoders instead of finite-state encoders.

In the blockwise coding scheme, a given sequence $x \in \mathcal{X}^\infty$ is encoded in the following manner. At first, x is parsed into non-overlapping blocks $x_{tk+1}^{(t+1)k}$ $(t = 0, 1, \cdots)$ of size k. Then, each block $x_{tk+1}^{(t+1)k}$ is encoded by $\psi_k : \mathcal{X}^k \to \mathcal{B}^*$, where ψ_k is called a *block encoder* of block size k, and the binary string $\psi_k(a_1^k)$ is called the *codeword* assigned to a_1^k for each $a_1^k \in \mathcal{X}^k$. A block encoder ψ_k is *uniquely decodable* if

$$\psi_k(a_1^k)\psi_k(a_{k+1}^{2k}) \cdots \psi_k(a_{(n-1)k+1}^{nk}) \neq \psi_k(\hat{a}_1^k)\psi_k(\hat{a}_{k+1}^{2k}) \cdots \psi_k(\hat{a}_{(m-1)k+1}^{mk})$$

for any $a_1^{nk} \in \mathcal{X}^{nk}$ and $\hat{a}_1^{mk} \in \mathcal{X}^{mk}$ such that $a_1^{nk} \neq \hat{a}_1^{mk}$ $(n, m \in \mathcal{N})$. It is clear that if ψ_k is uniquely decodable then there exists a corresponding decoder ψ_k^{-1} satisfying

$$a_1^k = \psi_k^{-1}\left(\psi_k(a_1^k)\right) \quad \forall a_1^k \in \mathcal{X}^k.$$

Moreover, a block encoder ψ_k is called a *prefix-free* or *instantaneous* if no codeword is a prefix of any other codeword. If ψ_k is prefix-free then ψ_k is uniquely decodable.

Now, we describe the well-known correspondence between a prefix-free encoder and a set of lengths of codewords.

Definition 2.4 (length function) Let A be a finite or countably infinite set. A function $\sigma : A \to \mathcal{N}$ is called a *length function* on A if σ satisfies the Kraft inequality:

$$\sum_{a \in A} 2^{-\sigma(a)} \leq 1. \tag{2.2}$$

Lemma 2.1 ([2]) For any uniquely decodable block encoder ψ_k, the function $\sigma(a_1^k) \equiv \ell(\psi_k(a_1^k))$ ($a_1^k \in \mathcal{X}^k$) is a length function on $\mathcal{X}^k$. On the other hand, if σ is a length function on $\mathcal{X}^k$, there exists a prefix-free block encoder ψ_k satisfying

$$\ell(\psi_k(a_1^k)) = \sigma(a_1^k) \quad \forall a_1^k \in \mathcal{X}^k.$$

Lemma 2.1 guarantees that we can regard a uniquely decodable block encoder as a length function without loss of generality. We will say "encode $a_1^k \in \mathcal{X}^k$ by a length function σ_k on $\mathcal{X}^k$" meaning that "encode a_1^k by a prefix-free encoder which encodes a_1^k using $\sigma_k(a_1^k)$ bits".

In this section, we investigate the optimal compression ratio attainable by blockwise encoders:

$$\rho^{bl}(x) \triangleq \inf_k \limsup_{T\to\infty} \inf_{\sigma_k} \frac{1}{Tk} \sum_{t=0}^{T-1} \sigma_k\left(x_{tk+1}^{(t+1)k}\right),$$

where $\inf_{\sigma_k}$ is taken over all length functions on $\mathcal{X}^k$. To characterize $\rho^{bl}(x)$, we introduce four kinds of quantities. The first two quantities are the overlapping empirical entropy $\hat{H}(x)$ and the Lempel-Ziv complexity $C(x)$. It should be noted that the definitions of $\hat{H}(x)$ and $C(x)$ of $x \in \mathcal{X}^\infty$ (Definition 2.2 and Definition 2.3) are valid even if the source alphabet $\mathcal{X}$ is countably infinite. In addition to these two quantities, we introduce the non-overlapping empirical entropy $\tilde{H}(x)$ and the self-entropy rate $h(x)$.

The non-overlapping empirical entropy $\tilde{H}(x)$ is a variation of the overlapping empirical entropy $\hat{H}(x)$.

Definition 2.5 (Non-overlapping empirical entropy) For each sequence $x \in \mathcal{X}^\infty$, the *non-overlapping empirical distribution* $q_k(\cdot|x_1^n)$ is defined as

$$q_k(a_1^k|x_1^n) \triangleq \frac{|\{t : 0 \le t \le \lfloor n/k \rfloor - 1, x_{tk+1}^{(t+1)k} = a_1^k\}|}{\lfloor n/k \rfloor} \quad \forall a_1^k \in \mathcal{X}^k. \tag{2.3}$$

Let

$$\tilde{H}_k(x_1^n) \triangleq -\frac{1}{k} \sum_{a_1^k \in \mathcal{X}^k} q_k(a_1^k|x_1^n) \log q_k(a_1^k|x_1^n),$$

and

$$\tilde{H}_k(x) \triangleq \limsup_{n\to\infty} \tilde{H}_k(x_1^n).$$

Then, the *non-overlapping empirical entropy* $\tilde{H}(x)$ of $x \in \mathcal{X}^\infty$ is defined as

$$\tilde{H}(x) \triangleq \inf_k \tilde{H}_k(x).$$

Next, we introduce the self-entropy rate $h(x)$ of a sequence $x \in \mathcal{X}^\infty$. For a source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ with alphabet $\mathcal{X}$ and a given string $x_1^n \in \mathcal{X}^n$, the quantity

$$-\frac{1}{n}\log P_{X_1^n}(x_1^n)$$

is called the *self-entropy* of x_1^n with respect to $\boldsymbol{X}$. The *self-entropy rate* of a sequence $x \in \mathcal{X}^\infty$ is defined as the asymptotic value of the infimum of the self-entropy over all the finite-state Markov sources.

Definition 2.6 (Self-entropy rate) For all $x \in \mathcal{X}^\infty$, define the *self-entropy rate* of x as

$$h(x) \triangleq \lim_{k\to\infty} \limsup_{n\to\infty} \inf_{\boldsymbol{X}\in\mathcal{P}_k} -\frac{1}{n}\log P_{X_1^n}(x_1^n),$$

where $\mathcal{P}_k$ denotes the set of all k-th order Markov sources with alphabet $\mathcal{X}$, and the initial state $x_{1-k}^0 \in \mathcal{X}^k$ of the source is a part of the specification of the source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty \in \mathcal{P}_k$.

Remark 2.1 Uyematsu and Kanaya [27] considered the quantity

$$h^{UK}(x) \triangleq \lim_{k\to\infty} \inf_{\boldsymbol{X}\in\mathcal{P}_k} \limsup_{n\to\infty} -\frac{1}{n}\log P_{X_1^n}(x_1^n),$$

for a sequence x over finite alphabet. According to the definitions of h and h^{UK}, it is apparent that $h(x) \le h^{UK}(x)$ for all $x \in \mathcal{X}^\infty$.

Before stating our main result, we define the finitely encodable sequences.

Definition 2.7 A sequence $x \in \mathcal{X}^\infty$ is *finitely encodable* (f.e.) if there exists a length function σ on $\mathcal{X}$ satisfying

$$\lim_{u\to\infty} \limsup_{n\to\infty} \sum_{a:\sigma(a)\ge u} p_1(a|x_1^n)\sigma(a) = 0. \tag{2.4}$$

Likewise, let $\mathcal{X}^\infty(\sigma)$ be the subset of $\mathcal{X}^\infty$ defined as

$$\mathcal{X}^\infty(\sigma) \triangleq \{x \in \mathcal{X}^\infty : x \text{ and } \sigma \text{ satisfy (2.4)}\}.$$

Now, we state our main result.

Theorem 2.2 For any f.e. sequence $x \in \mathcal{X}^\infty$, we have

$$\hat{H}(x) = \lim_{k\to\infty} \hat{H}_k(x), \quad \tilde{H}(x) = \lim_{k\to\infty} \tilde{H}_k(x), \tag{2.5}$$

and

$$C(x) \le h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x). \tag{2.6}$$

Remark 2.2 Since every sequence x over a finite alphabet $\mathcal{X}$ is f.e., Theorem 2.2 contains finite alphabet case as a special case. In finite alphabet case, $C(x) \leq \hat{H}(x)$ and $C(x) \leq h(x)$ have been known ([36], [22]). $\hat{H}(x) = \lim_{k\to\infty} \hat{H}_k(x)$ has been shown in [36, Lemma 1]. Moreover, for any stationary ergodic source with finite alphabet, $\hat{H}(x) = \tilde{H}(x)$ holds for almost all x [25, Remark II.3.6].

2.5 Proof of Theorem 2.2

In this section, we prove Theorem 2.2. Before proving the theorem, we prepare some lemmas.

At first, we show a relation between a length function and a probability distribution.

Lemma 2.2 Let μ and $\hat{\mu}$ be non-negative real-valued functions on A. If $\sum_{a\in A}\mu(a) = 1$ and $\sum_{a\in A}\hat{\mu}(a) \leq 1$ then

$$-\sum_{a\in A}\mu(a)\log\mu(a) \leq -\sum_{a\in A}\mu(a)\log\hat{\mu}(a). \tag{2.7}$$

Especially, for any length function σ on A,

$$-\sum_{a\in A}\mu(a)\log\mu(a) \leq \sum_{a\in A}\mu(a)\sigma(a). \tag{2.8}$$

Proof: The inequality (2.7) is well known as the non-negativity of the divergence (see [2, 9, 7]). Letting $\hat{\mu}(a) = 2^{-\sigma(a)}$, we have (2.8). □

Next, we state some properties of f.e. sequences. The following lemma shows that if $x \in \mathcal{X}^\infty(\sigma)$ then σ encodes x with finite rate.

Lemma 2.3 For any length function σ on $\mathcal{X}$ and any sequence $x \in \mathcal{X}^\infty(\sigma)$, we have

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma(x_i) < \infty. \tag{2.9}$$

Proof: If $x \in \mathcal{X}^\infty(\sigma)$, then for any $\gamma > 0$ and for sufficiently large u and n, we have

$$\frac{1}{n}\sum_{i=1}^{n}\sigma(x_i) \leq \frac{1}{n}\sum_{i=1}^{n}u + \sum_{a:\sigma(a)\geq u} p_1(a|x_1^n)\sigma(a) \leq u + \gamma$$

and thus, we have (2.9). □

The following lemma shows that if $x \in \mathcal{X}^\infty(\sigma)$ then sufficiently small fraction of symbols in x does not affect the rate.

Lemma 2.4 Let σ be a length function on $\mathcal{X}$ and assume that $x \in \mathcal{X}^\infty(\sigma)$. For any $\gamma > 0$, there exists $\epsilon > 0$ and $n_0 \in \mathcal{N}$ such that for any $n \geq n_0$ and any set $B \subset \{1, 2, \cdots, n\}$ satisfying $|B|/n < \epsilon$,

$$\frac{1}{n}\sum_{j \in B} \sigma(x_j) \leq \gamma.$$

Proof: Since $x \in \mathcal{X}^\infty(\sigma)$, for any $\gamma > 0$, we can choose u and ϵ so that

$$\limsup_{n\to\infty} \sum_{a:\sigma(a)\geq u} p_1(a|x_1^n)\sigma(a) \leq \gamma/4.$$

Fix $\epsilon > 0$ so that $\epsilon u \leq \gamma/2$. Then, for any $B \subset \{1, 2, \cdots, n\}$ satisfying $|B| \leq \epsilon n$, we have

$$\begin{aligned} \frac{1}{n}\sum_{j\in B}\sigma(x_j) &= \frac{1}{n}\sum_{\substack{j\in B:\\ \sigma(x_j)<u}}\sigma(x_j) + \frac{1}{n}\sum_{\substack{j\in B:\\ \sigma(x_j)\geq u}}\sigma(x_j) \\ &\leq \epsilon u + (\gamma/4 + \gamma/4) \\ &\leq \gamma, \end{aligned}$$

for sufficiently large n. Hence, the lemma is proved. □

It should be noted that the combination of (2.8) and (2.9) indicates that $\hat{H}_1(x) < \infty$ and $\tilde{H}_1(x) < \infty$ for any f.e. sequence x. Thus, if x is f.e., we have $\hat{H}(x) = \inf_k \hat{H}_k(x) < \infty$, and $\tilde{H}(x) = \inf_k \tilde{H}_k(x) < \infty$.

Next, we show an inequality which provides a connection between a parsing of a string and the self-entropy.

Lemma 2.5 Let $x \in \mathcal{X}^\infty$ be an f.e. sequence and $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ be a k-th order Markov source with the initial state $x_{1-k}^0 \in \mathcal{X}^k$. For any $n \in \mathcal{N}$ and arbitrary parsing $x_1^n = w_1 w_2 \cdots w_t$ of x_1^n ($w_i \in \bigcup_l \mathcal{X}^l$), we have

$$\frac{1}{n}\sum_{i=1}^{t} \log\frac{t}{\eta(w_i)} \leq -\frac{1}{n}\log P_{X_1^n}(x_1^n) + \varepsilon(t/n), \tag{2.10}$$

where $\eta(w) \triangleq |\{j : 1 \leq j \leq t, w_j = w\}|$, and $\varepsilon(t/n)$ is a constant which does not depend on $\boldsymbol{X}$ and satisfies that $\varepsilon(t/n) \to 0$ as $n \to \infty$ and $t/n \to 0$.

Proof: For any parsing $w_1, w_2, \cdots, w_t$, let n_i be the index of the beginning of the i-th phrase, i.e., $w_i = x_{n_i}^{n_{i+1}-1}$. For each $i = 1, \cdots, t$, let s_i be the k symbols of x preceding w_i, that is, $s_i = x_{n_i-k}^{n_i-1}$. Especially, $s_1 = x_{1-k}^0$. For each $i = 1, \cdots, t$, let l_i be the length of w_i, that is, $l_i = n_{i+1} - n_i$. For $w \in \mathcal{X}^*$, $s \in \mathcal{X}^k$ and integer l, define

$$\begin{aligned}
\eta(w, s) &\triangleq |\{j : 1 \le j \le t, w_j = w \text{ and } s_j = s\}|, \\
\eta(l, s) &\triangleq |\{j : 1 \le j \le t, s_j = s \text{ and } l_j = l\}|, \\
\eta(l) &\triangleq |\{j : 1 \le j \le t, l_j = l\}|.
\end{aligned}$$

Using η, define random variables W, S, L as

$$\Pr\{W = w, S = s, L = l\} = \begin{cases} \eta(w, s)/t, & \text{if } l \text{ equals to the length of } w \\ 0, & \text{otherwise.} \end{cases}$$

Since the source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ is a k-th order Markov source with the initial state $x_{1-k}^0 \in \mathcal{X}^k$, we have

$$\begin{aligned}
-\log P_{X_1^n}(x_1^n) &= -\sum_{i=1}^t \log P_{X_{k+1}|X_1^k}(w_i|s_i) \\
&= -\sum_{l,s} \sum_{i:l_i=l,s_i=s} \log P_{X_{k+1}|X_1^k}(w_i|s_i) \\
&= -t \sum_{l,s} \frac{\eta(l, s)}{t} \sum_{w \in \mathcal{X}^l} \frac{\eta(w, s)}{\eta(l, s)} \log P_{X_{k+1}|X_1^k}(w|s) \\
&= -t \sum_{l,s} \Pr\{S = s, L = l\} \\
&\quad \times \sum_{w \in \mathcal{X}^l} \Pr\{W = w|S = s, L = l\} \log P_{X_{k+1}|X_1^k}(w|s) \\
&\overset{\text{(a)}}{\ge} -t \sum_{l,s} \Pr\{S = s, L = l\} \\
&\quad \times \Bigg\{ \sum_{w \in \mathcal{X}^l} \Pr\{W = w|S = s, L = l\} \\
&\quad \times \log \Pr\{W = w|S = s, L = l\} \Bigg\} \\
&= tH(W|S, L) \\
&\ge tH(W) - tH(L) - tH(S) \\
&= t \sum_{w \in \mathcal{X}^*} \frac{\eta(w)}{t} \log \frac{t}{\eta(w)} - tH(L) - tH(S)
\end{aligned}$$

$$= \sum_{i=1}^{t} \log \frac{t}{\eta(w_i)} - tH(L) - tH(S),$$

where the inequality (a) is due to Lemma 2.2. Hence, we have

$$\frac{1}{n}\sum_{i=1}^{t} \log \frac{t}{\eta(w_i)} \le -\frac{1}{n} \log P_{X_1^n}(x_1^n) + \frac{t}{n} H(L) + \frac{t}{n} H(S). \tag{2.11}$$

Now, we bound the second term $(t/n)H(L)$ of the right-hand side of (2.11). Since the expectation EL of L satisfies $EL = n/t$, according to [9, corollary 2.4], we have

$$\frac{t}{n} H(L) \le \frac{t}{n}[\log(EL) + \log e] = \frac{t}{n}\log(n/t) + \frac{t}{n}\log e.$$

So, $(t/n)H(L) \to 0$ as $t/n \to 0$. Next, we upper bound the third term $(t/n)H(S)$ of the right-hand side of (2.11). Let σ be the length function satisfying (2.4). By using σ and the initial state $x_{1-k}^0 \in \mathcal{X}^k$, we consider the following encoder. If the symbol $a \in \mathcal{X}$ appeared in the initial state $x_{1-k} \cdots x_0$, we encode the symbol a into its index i $(0 \le i \le k-1)$ such that $a = x_{-i}$. Otherwise, we encode $a \in \mathcal{X}$ by using a code corresponding to the length function σ. This implies the length function $\tilde{\sigma}$ defined by

$$\tilde{\sigma}(a) = \begin{cases} 1 + \lceil \log k \rceil & \text{if } a = x_{-i} \text{ for some } i = 0, \cdots, k-1 \\ 1 + \sigma(a) & \text{otherwise.} \end{cases}$$

Let B be the set of index defined by

$$B \triangleq \bigcup_{i: n_i > k} \{n_i - k, n_i - k + 1, \cdots, n_i - 1\}.$$

Then, considering the encoding of the sequence $x_{n_i-k}^{n_i-1}$ into a codeword of length $\sum_{j=1}^{k} \tilde{\sigma}(x_{n_i-j})$ for $i = 1, 2, \cdots, t$, we have

$$\begin{aligned} tH(S) &\le k\left\{ \sum_{-(k-1)\le i \le 0} \tilde{\sigma}(x_i) + \sum_{i\in B} \tilde{\sigma}(x_i) \right\} \\ &\le k^2(1 + \lceil \log k \rceil) + k \sum_{i \in B} \left\{1 + \lceil \log k \rceil + \sigma(x_i)\right\}. \end{aligned}$$

Since $|B|/n \le kt/n \to 0$ as $t/n \to 0$, according to Lemma 2.4,

$$\frac{t}{n} H(S) \le \frac{k^2(\log k + 2)}{n} + \frac{k|B|(\log k + 2)}{n} + \frac{k}{n}\sum_{i\in B}\sigma(x_i) \to 0,$$

as $n \to \infty$ and $t/n \to 0$. According to the above discussion, we have

$$\frac{1}{n}\sum_{i=1}^{t}\log\frac{t}{\eta(w_i)} \le -\frac{1}{n}\log P_{X_1^n}(x_1^n) + \varepsilon(t/n),$$

where $\varepsilon(t/n) \to 0$ as $t/n \to 0$. □

The following lemma shows that the maximum number $c(x_1^n)$ of phrases in arbitrary distinct parsing of x_1^n divided by n tends to zero if x is f.e.

Lemma 2.6 For any f.e. sequence $x \in \mathcal{X}^\infty$, we have

$$\lim_{n\to\infty}\frac{c(x_1^n)}{n} = 0. \tag{2.12}$$

Proof: Since x is f.e., there exists a length function σ on $\mathcal{X}$ which satisfies (2.9), and thus, there exists a prefix-free encoder $\psi : \mathcal{X} \to \mathcal{B}^*$ satisfying $\ell(\psi(a)) = \sigma(a)$ $(a \in \mathcal{X})$. Let

$$x_1^n = x_{n_0+1}^{n_1} x_{n_1+1}^{n_2} x_{n_2+1}^{n_3} \cdots x_{n_{c-1}+1}^{n_c}$$

be a parsing of x_1^n into $c = c(x_1^n)$ distinct phrases, where $n_0 = 0$ and $n_c = n$. For each $i = 1, 2, \cdots, c$, let

$$u_i = \psi(x_{n_{i-1}+1})\psi(x_{n_{i-1}+2})\cdots\psi(x_{n_i}).$$

Since ψ if prefix-free and phrases $x_{n_{i-1}+1}^{n_i}$ $(i = 1, \cdots, c)$ are distinct, all $u_1, u_2, \cdots, u_c$ are also distinct, and thus,

$$|\{i : \ell(u_i) = l\}| \le 2^l.$$

Let k and q be integers such that

$$c = \sum_{l=0}^{k} 2^l + q = 2^{k+1} - 1 + q, \tag{2.13}$$

and

$$0 \le q < 2^{k+1}. \tag{2.14}$$

Then, we can underestimate $\sum_{i=1}^{n}\sigma(x_i)$ as

$$\begin{aligned}\sum_{i=1}^{n}\sigma(x_i) &= \sum_{i=1}^{c}\ell(u_i) \\ &\ge \sum_{l=0}^{k} l2^l + (k+1)q \\ &= (k-1)2^{k+1} + 2 + (k+1)q \\ &= (k-1)(c+1) + 2(q+1). \end{aligned} \tag{2.15}$$

From (2.13), we have

$$k - 1 = \log(c - q + 1) - 2,$$

which together with (2.15) yields

$$\begin{aligned}
\sum_{i=1}^{n} \sigma(x_i) &\geq \{\log(c-q+1) - 2\}(c+1) + 2(q+1) \\
&= (c+1)\left\{\log\frac{c+1}{4} + \frac{2(q+1)}{c+1} - \log\left(1 + \frac{q}{c-q+1}\right)\right\} \\
&= (c+1)\left\{\log\frac{c+1}{4} + \frac{2}{c+1} + \frac{2\tau}{1+\tau} - \log(1+\tau)\right\}, \qquad (2.16)
\end{aligned}$$

where

$$\tau = \frac{q}{c-q+1}.$$

From (2.13) and (2.14), $0 \leq \tau < 1$, and thus, $2\tau/(1+\tau) - \log(1+\tau) > 0$. Hence, (2.16) yields

$$\sum_{i=1}^{n} \sigma(x_i) \geq (c+1)\log\frac{c+1}{4} + 2,$$

and thus,

$$\frac{c}{n} \leq \frac{(1/n)\sum_{i=1}^{n}\sigma(x_i)}{\log(c/4)}.$$

From (2.9), we have $\limsup_{n\to\infty}(1/n)\sum_{i=1}^{n}\sigma(x_i) < \infty$, while $\log(c/4) \to \infty$ as $n \to \infty$. Hence, we have the lemma. □

The following two lemmas imply that $C(x) \leq h(x)$ and $\tilde{H}(x) \leq h(x)$ if x is f.e.

Lemma 2.7 For any f.e. sequence $x \in \mathcal{X}^\infty$, we have

$$C(x) \leq h(x).$$

Proof: Fix $\epsilon > 0$ arbitrary. Choose $k \in \mathcal{N}$, sufficiently large $n \in \mathcal{N}$ and $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty} \in \mathcal{P}_k$ so that,

$$-\frac{1}{n}\log P_{X_1^n}(x_1^n) \leq h(x) + \epsilon/2. \qquad (2.17)$$

On the other hand, for any distinct parsing $x_1^n = w_1 \cdots w_c$ ($c \equiv c(x_1^n)$) of x_1^n, we have $\eta(w_i) = 1$ ($i = 1, \cdots, c$). Thus, we have

$$\sum_{i=1}^{c}\log\frac{c}{\eta(w_i)} = c\log c = c\log n - c\log(n/c).$$

By Lemma 2.5, (2.17) and (2.12), we have

$$\begin{aligned}\frac{1}{n}c\log n &\leq -\frac{1}{n}\log P_{X_1^n}(x_1^n)+\epsilon/2,\\ &\leq h(x)+\epsilon,\end{aligned}$$

for sufficiently large n. Since ϵ can be chosen arbitrarily small, we have

$$\limsup_{n\to\infty}\frac{c(x_1^n)\log n}{n}\leq h(x).$$

Hence, we have the lemma. □

Lemma 2.8 Let $x\in\mathcal{X}^\infty$ be an f.e. sequence. For any $l\in\mathcal{N}$ and any $\gamma>0$, there exists an integer $k_0=k_0(x,l,\gamma)$ such that for any $k\geq k_0$,

$$\tilde{H}_k(x)\leq\limsup_{n\to\infty}\inf_{\boldsymbol{X}\in\mathcal{P}_l}-\frac{1}{n}\log P_{X_1^n}(x_1^n)+\gamma.$$

Proof: Fix $m\in\mathcal{N}$. Parse x_1^{mk} into blocks of block size k, that is,

$$x_1^{mk}=x_1^k x_{k+1}^{2k}\cdots x_{(t-k)k+1}^{tk}=w_1w_2\cdots w_t,$$

where $w_i=x_{(i-1)k+1}^{ik}$ for each $i=1,\cdots,t$. Apply Lemma 2.5 to this parsing $x_1^{mk}=w_1\cdots w_t$, then, for any $\boldsymbol{X}=\{X_i\}_{i=1}^\infty\in\mathcal{P}_l$, we have

$$\tilde{H}(x_1^{mk})\leq-\frac{1}{mk}\log P_{X_1^{mk}}(x_1^{mk})+\varepsilon(1/k).$$

Thus, we have

$$\begin{aligned}\tilde{H}_k(x) &= \limsup_{n\to\infty}\tilde{H}_k(x_1^n)\\ &= \limsup_{m\to\infty}\tilde{H}_k(x_1^{mk})\\ &\leq \limsup_{m\to\infty}\inf_{\boldsymbol{X}\in\mathcal{P}_l}-\frac{1}{mk}\log P_{X_1^{mk}}(x_1^{mk})+\varepsilon(1/k)+\gamma/2\\ &\leq \limsup_{n\to\infty}\inf_{\boldsymbol{X}\in\mathcal{P}_l}-\frac{1}{n}\log P_{X_1^{mk}}(x_1^n)+\varepsilon(1/k)+\gamma/2.\end{aligned}$$

Since $\varepsilon(1/k)<\gamma/2$ for sufficiently large k, we obtain the lemma. □

Now, we introduce the cyclic version of the overlapping empirical entropy:

$$H_k^\circ(x_1^n)\triangleq-\frac{1}{k}\sum_{a_1^k\in\mathcal{X}^k}p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n),$$

where

$$p_k^\circ(a_1^k|x_1^n) \triangleq \frac{|\{i : 0 \le i < n, \breve{x}_{i+1}^{i+k} = a_1^k\}|}{n},$$

and $\breve{x}_1^n \in \mathcal{X}^{n+k-1}$ is defined as $\breve{x}_1^n = x_1^n$ and $\breve{x}_{n+1}^{n+k-1} = x_1^{k-1}$. Next lemma shows that $H_k^\circ(x_1^n)$ is asymptotically equal to $\hat{H}_k(x_1^n)$.

Lemma 2.9 For any f.e. sequence $x \in \mathcal{X}^\infty$, we have

$$\left|H_k^\circ(x_1^n) - \hat{H}_k(x_1^n)\right| \to 0 \ (n \to \infty). \tag{2.18}$$

Proof: Let $B_k \subset \mathcal{X}^k$ be the set defined as

$$B_k \triangleq \{\breve{x}_i^{i+k-1} : i = n-k+2, n-k+3, \cdots, n\}.$$

Then, we have

$$\begin{aligned}
\left|H_k^\circ(x_1^n) - \hat{H}_k(x_1^n)\right| &= \Bigg|-\frac{1}{k}\sum_{a_1^k\in\mathcal{X}^k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n) \\
&\quad +\frac{1}{k}\sum_{a_1^k\in\mathcal{X}^k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n)\Bigg| \\
&\le \Bigg|-\frac{1}{k}\sum_{a_1^k\in B_k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n) \\
&\quad +\frac{1}{k}\sum_{a_1^k\in B_k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n)\Bigg| \\
&\quad +\Bigg|-\frac{1}{k}\sum_{a_1^k\notin B_k} p_k^\circ(a_1^k|x_1^n)\log p_k^\circ(a_1^k|x_1^n) \\
&\quad +\frac{1}{k}\sum_{a_1^k\notin B_k} p_k(a_1^k|x_1^n)\log p_k(a_1^k|x_1^n)\Bigg|.
\end{aligned} \tag{2.19}$$

The first term of (2.19) is upper-bounded as follows. The function $f(t) \triangleq -t\log t$ satisfies

$$|f(t) - f(t+\tau)| \le -\tau\log\tau,$$

for $0 \le \tau \le 1/2$ and $0 \le t \le 1-\tau$ (see [3, the proof of Lemma 2.7]). Since

$$|p_k^\circ(a_1^k|x_1^n) - p_k(a_1^k|x_1^n)| \le \frac{2k}{n-k+1}, \quad \forall a_1^k \in B_k,$$

we can choose sufficiently large n such that $\tau \equiv |p_k^\circ(a_1^k|x_1^n) - p_k(a_1^k|x_1^n)|$ satisfies $0 \le \tau \le 1/2$ and $t \equiv \min\{p_k^\circ(a_1^k|x_1^n), p_k(a_1^k|x_1^n)\}$ satisfies $0 \le t \le 1-\tau$ for each $a_1^k \in B_k$. Hence, we have

$$\begin{aligned}
&\left| -\frac{1}{k} \sum_{a_1^k \in B_k} p_k^\circ(a_1^k|x_1^n) \log p_k^\circ(a_1^k|x_1^n) + \frac{1}{k} \sum_{a_1^k \in B_k} p_k(a_1^k|x_1^n) \log p_k(a_1^k|x_1^n) \right| \\
&\le \frac{1}{k} \sum_{a_1^k \in B_k} \left| f\left(p_k^\circ(a_1^k|x_1^n)\right) - f\left(p_k(a_1^k|x_1^n)\right) \right| \\
&\le -\frac{1}{k} \sum_{a_1^k \in B_k} \frac{2k}{n-k+1} \log \frac{2k}{n-k+1} \\
&\le -\frac{2k}{n-k+1} \log \frac{2k}{n-k+1} \to 0 \quad (n \to \infty),
\end{aligned}$$

where the last inequality follows from $|B_k| \le k$.

On the other hand, the second term of (2.19) is upper-bounded as follows. By the definition of p_k°,

$$p_k^\circ(a_1^k|x_1^n) = \frac{n-k+1}{n} p_k(a_1^k|x_1^n), \quad \forall a_1^k \notin B_k,$$

and thus

$$\begin{aligned}
-\sum_{a_1^k \notin B_k} p_k^\circ(a_1^k|x_1^n) \log p_k^\circ(a_1^k|x_1^n) &= -\frac{n-k+1}{n} \sum_{a_1^k \notin B_k} p_k(a_1^k|x_1^n) \log p_k(a_1^k|x_1^n) \\
&\quad -\frac{n-k+1}{n} \sum_{a_1^k \notin B_k} p_k(a_1^k|x_1^n) \log \frac{n-k+1}{n}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\left| -\frac{1}{k} \sum_{a_1^k \notin B_k} p_k^\circ(a_1^k|x_1^n) \log p_k^\circ(a_1^k|x_1^n) + \frac{1}{k} \sum_{a_1^k \notin B_k} p_k(a_1^k|x_1^n) \log p_k(a_1^k|x_1^n) \right| \\
&\le \frac{k-1}{n} \hat{H}_k(x_1^n) - \frac{n-k+1}{nk} \log \frac{n-k+1}{n}.
\end{aligned}$$

Since x is f.e., $\hat{H}_k(x) < \infty$, and thus the second term of (2.19) tends to 0 as $n \to \infty$. This completes the proof of (2.18). □

Lemma 2.10 For any f.e. sequence $x \in \mathcal{X}^\infty$ and any integer $k \ge 1$, we have

$$\limsup_{n\to\infty} \inf_{\boldsymbol{X} \in \mathcal{P}_{k-1}} -\frac{1}{n} \log P_{X_1^n}(x_1^n) \le \hat{H}_k(x).$$

Proof: Fix $n \in \mathcal{N}$. From $p_k^\circ(\cdot|x_1^n)$, we can define a set of random variables $X_1, X_2, \cdots, X_k$ such that (i) the distribution of X_1^k satisfies

$$\Pr\{X_1^k = a_1^k\} \stackrel{\triangle}{=} p_k^\circ(a_1^k|x_1^n) \quad \forall a_1^k \in \mathcal{X}^k,$$

and (ii) X_1^k is shift-invariant, that is, for all $i, l \in \mathcal{N}$ $(i+l-1 \leq k)$,

$$\Pr\{X_1^l = a_1^l\} = \Pr\{X_i^{i+l-1} = a_1^l\} \quad \forall a_1^k \in \mathcal{X}^k. \tag{2.20}$$

Then, we have

$$\begin{aligned}
H_k^\circ(x_1^n) &= \frac{1}{k}\sum_{a_1^k \in \mathcal{X}^k} p_k^\circ(a_1^k|x_1^n) \log \frac{1}{p_k^\circ(a_1^k|x_1^n)} \\
&= \frac{1}{k} H(X_1^k) \\
&= \frac{1}{k}\sum_{i=1}^k H(X_i|X_1, \cdots, X_{i-1}) && (2.21) \\
&= \frac{1}{k}\sum_{i=1}^k H(X_k|X_{k-i+1}, \cdots, X_{k-1}) && (2.22) \\
&\geq \frac{1}{k}\sum_{i=1}^k H(X_k|X_1, \cdots, X_{k-1}) && (2.23) \\
&= H(X_k|X_1, \cdots, X_{k-1}) && (2.24)
\end{aligned}$$

where (2.21) follows from the chain rule [2], (2.22) from (2.20), and (2.23) from the fact that conditioning reduces entropy [2].

On the other hand, let $\hat{\boldsymbol{X}} = \{\hat{X}_i\}_{i=1}^\infty$ be a $(k-1)$-th order Markov source such that

$$P_{\hat{X}_k|\hat{X}_1^{k-1}}(a_k|a_1^{k-1}) \stackrel{\triangle}{=} \frac{p_k^\circ(a_1^k|x_1^n)}{\sum_{a\in\mathcal{X}} p_k^\circ(a_1^{k-1}a|x_1^n)} \quad \forall a_1^k \in \mathcal{X}^k,$$

with the initial state x_{n-k}^n. Then, by the definitions, we have

$$\begin{aligned}
\inf_{\boldsymbol{X}\in\mathcal{P}_{k-1}} -\frac{1}{n}\log P_{X_1^n}(x_1^n) &\leq -\frac{1}{n}\log P_{\hat{X}_1^n}(x_1^n) \\
&= H(X_k|X_1^{k-1}) \\
&\leq H_k^\circ(x_1^n),
\end{aligned}$$

where the last inequality follows from (2.24). By using (2.18), we have

$$\limsup_{n\to\infty} \inf_{\boldsymbol{X}\in\mathcal{P}_{k-1}} -\frac{1}{n}\log P_{X_1^n}(x_1^n) \leq \hat{H}_k(x).$$

Hence, we have the lemma. □

Lemma 2.11 Let $x \in \mathcal{X}^\infty$ be an f.e. sequence. For any integer k and any $\epsilon > 0$, there exists an integer $m_0 = m_0(x, k, \epsilon)$ such that for all $m \geq m_0$,

$$\hat{H}_m(x) \leq \tilde{H}_k(x) + \epsilon.$$

Proof: Let n_0 be an integer such that

$$\tilde{H}_k(x_1^n) \leq \tilde{H}_k(x) + \epsilon/2, \quad \forall n \geq n_0.$$

Choose integers n and m satisfying $n \geq n_0$ and $n \geq m \geq k$. Let ψ_k be a function on $\mathcal{X}^k$ defined as

$$\psi_k(a_1^k) \stackrel{\triangle}{=} -\log q_k(a_1^k|x_1^n), \quad \forall a_1^k \in \mathcal{X}^k,$$

and σ be a length function on $\mathcal{X}$ which satisfies (2.4). Using ψ_k and σ, define a function $\phi_{m,j}$ $(j = 0, 1, \cdots, k-1)$ over $\mathcal{X}^m$ by

$$\phi_{m,j}(a_1^m) \stackrel{\triangle}{=} \sum_{i=1}^{j} \sigma(a_i) + \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} \psi_k(a_{i+1}^{i+k}) + \sum_{i=r+1}^{m} \sigma(a_i),$$

where $r = \max\{i : i \equiv j \bmod k, 1 \leq i \leq m\}$. Figure 2.1 shows how to construct $\phi_{m,j}$ from ψ_k and σ. Moreover, using $\phi_{m,j}$ $(j = 0, \cdots, k-1)$, define a function ϕ_m^* on $\mathcal{X}^m$ by

$$\phi_m^*(a_1^m) \stackrel{\triangle}{=} \min_{j: 0 \leq j \leq k-1} \phi_{m,j}(a_1^m) + \lceil \log k \rceil. \tag{2.25}$$

Then, we have

$$\begin{aligned} &(n-m+1)m\hat{H}_m(x_1^n) \\ &\quad = \; -(n-m+1) \sum_{a_1^m \in \mathcal{X}^m} p_m(a_1^m|x_1^n) \log p_m(a_1^m|x_1^n), \\ &\quad \leq \; \sum_{i=1}^{n-m+1} \phi_m^*(x_i^{i+m-1}), \end{aligned} \tag{2.26}$$

where the inequality is due to Lemma 2.2. According to (2.25), we have

$$\begin{aligned} \sum_{i=1}^{n-m+1} \phi_m^*(x_i^{i+m-1}) &\leq \sum_{i=1}^{n-m+1} \phi_{m,j_i}(x_i^{i+m-1}) + (n-m+1)\lceil \log k \rceil \\ &\leq \sum_{i=1}^{n-m+1} \phi_{m,j_i}(x_i^{i+m-1}) + n(\log k + 1), \end{aligned} \tag{2.27}$$

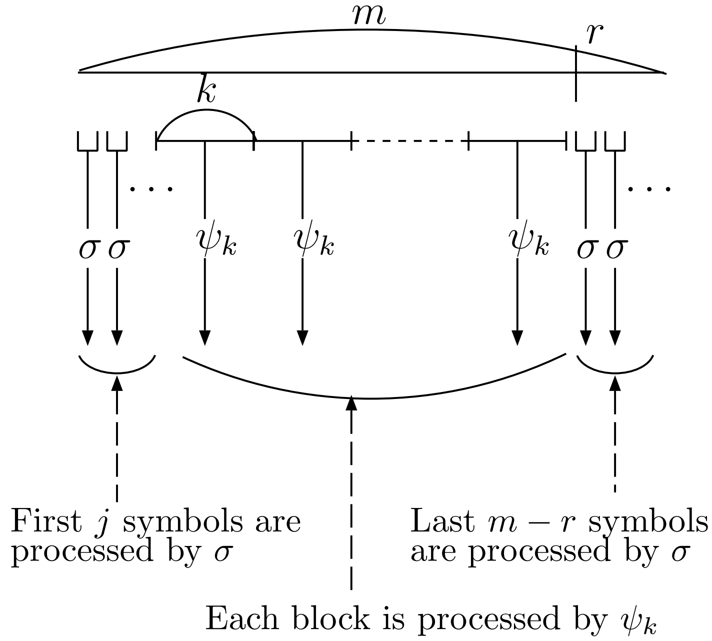

Figure 2.1: The construction of $\phi_{m,j}$

where j_i denotes the integer satisfying $0 \le j_i \le k-1$ and $i+j_i \equiv 1 \bmod k$. To evaluate the first term of the right-hand side of (2.27), recall how each block x_i^{i+m-1} is processed by ϕ_{m,j_i} (see Figure 2.1). By the construction of ϕ_{m,j_i}, (i) the first j_i symbols of x_i^{i+m-1} and the last $m-r$ ($r = \max\{r' : r' \equiv j_i \bmod k, 1 \le r' \le m\}$) symbols of x_i^{i+m-1} are processed by σ, and (ii) each blocks $x_{i+j_i+tk}^{i+j_i+(t+1)k}$ ($t = 0, 1, \cdots$) of the length k in x_{i+1}^{i+m} are processed by ψ_k. Put them all together, each non-overlapping block $x_{tk+1}^{(t+1)k}$ ($t = 0, 1, \cdots \lfloor n/k \rfloor - 1$) of block size k in x_1^n is processed by ψ_k at most $m-k$ times, and each symbol in x_1^n is processed by σ at most $k-1$ times (see Fig. 2.2). Hence, (2.27) can be upper-bounded by

$$\begin{aligned}
&\sum_{i=1}^{n-m+1} \phi_{m,j_i}(x_i^{i+m-1}) + n(\log k + 1) \\
\le\ & (m-k) \sum_{t=0}^{\lfloor n/k \rfloor - 1} \psi_k\left(x_{tk+1}^{(t+1)k}\right) + k \sum_{i=1}^{n} \sigma(x_i) + n(\log k + 1) \\
\le\ & m \left\lfloor \frac{n}{k} \right\rfloor \left(- \sum_{a_1^k \in \mathcal{X}^k} q_k(a_1^k | x_1^n) \log q_k(a_1^k | x_1^n), \right) + mn\varepsilon
\end{aligned}$$

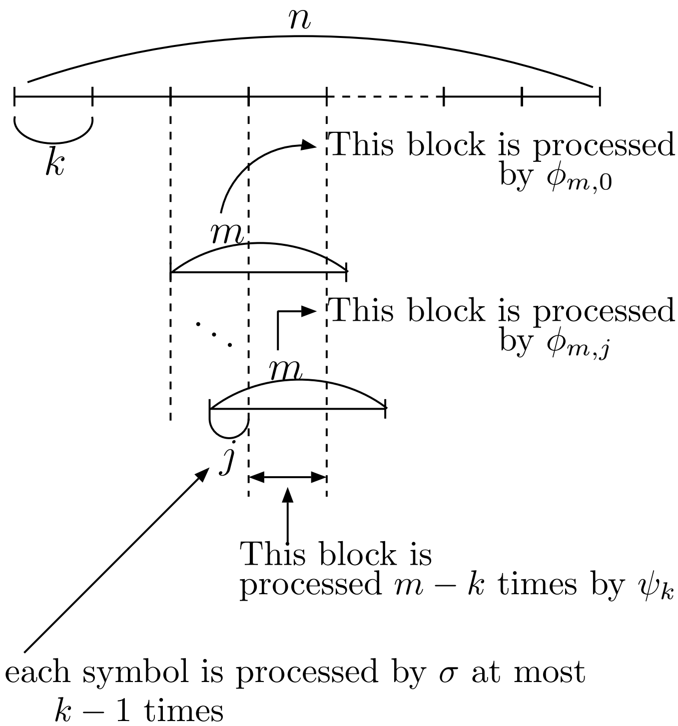

Figure 2.2: Calculating $\sum_{i=1}^{n-m+1} \phi_{m,j_i}(x_i^{i+m-1})$

$$\leq \quad mn\tilde{H}_k(x_1^n) + mn\varepsilon, \tag{2.28}$$

where ε is defined as

$$\varepsilon \triangleq \frac{\log k + 1}{m} + \frac{k}{m}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma(x_i)\right).$$

Combination of (2.26), (2.27), and (2.28) yields

$$\hat{H}_m(x) \leq \frac{n}{n-m+1}\left\{\tilde{H}_k(x) + \varepsilon\right\}.$$

Since x is f.e., $(1/n)\sum_i \sigma(x_i) < \infty$ and thus, $\varepsilon < \epsilon/2$ if $m \gg k$. Hence, if n and m are sufficiently large, we have

$$\hat{H}_m(x) \leq \tilde{H}_k(x) + \epsilon,$$

and thus, lemma is proved. □

Now, we prove our main result.

Proof of Theorem 2.2: According to Lemma 2.8–2.11, it can be shown that for any k and any $\epsilon > 0$, there exists an integer M such that for any $m \geq M$, $\tilde{H}_m(x) \leq \tilde{H}_k(x) + \epsilon$. Thus, for any k,

$$\tilde{H}(x) \leq \liminf_{m\to\infty} \tilde{H}_m(x) \leq \limsup_{m\to\infty} \tilde{H}_m(x) \leq \tilde{H}_k(x).$$

Since, for any $\epsilon > 0$, we can choose k such that $\tilde{H}_k(x) \leq \tilde{H}(x) + \epsilon$, we have $\tilde{H}(x) = \lim_{k\to\infty} \tilde{H}_k(x)$. Similarly, it can be shown that $\hat{H}(x) = \lim_{k\to\infty} \hat{H}_k(x)$. Hence, (2.5) is proved.

Next, we prove (2.6). Lemma 2.7 shows that

$$C(x) \leq h(x),$$

and thus, the leftmost inequality in (2.6) is proved. Moreover, by Lemma 2.8, Lemma 2.10, and Lemma 2.11, we have

$$\tilde{H}(x) \leq h(x) \leq \hat{H}(x) \leq \tilde{H}(x),$$

and thus,

$$\tilde{H}(x) = h(x) = \hat{H}(x).$$

Lastly, we prove $\tilde{H}(x) = \rho^{bl}(x)$. We can construct the Shannon code [2, 7] $\bar{\sigma}_k$ for the probability distribution $q_k(\cdot|x_1^{Tk})$ which satisfies that

$$E_{q_k}\left[\bar{\sigma}_k(X_1^k)\right] \leq k\tilde{H}(x_1^{Tk}) + 2,$$

where E_{q_k} denotes the expectation with respect to $q_k(\cdot|x_1^{Tk})$. Hence, we have

$$\frac{1}{Tk}\sum_{t=0}^{T-1}\bar{\sigma}_k\left(x_{tk+1}^{(t+1)k}\right) \le \tilde{H}(x_1^{Tk}) + (2/k),$$

and thus,

$$\rho^{bl}(x) \le \tilde{H}(x).$$

On the other hand, by Lemma 2.2, we have

$$\tilde{H}(x_1^{Tk}) \le \inf_{\sigma_k}\frac{1}{Tk}\sum_{t=0}^{T-1}\sigma_k\left(x_{tk+1}^{(t+1)k}\right),$$

and thus

$$\tilde{H}(x) \le \rho^{bl}(x).$$

Hence, we have $\tilde{H}(x) = \rho^{bl}(x)$, and thus, (2.6) is proved. □

2.6 Applications to Coding Theorems

Theorem 2.2 shows that the optimal performance $\rho^{bl}(x)$ attainable by block-wise coding is equal to $h(x)$, $\hat{H}(x)$, and $\tilde{H}(x)$. Moreover, it is also shown that $\rho^{bl}(x)$ is lower bounded by $C(x) = \limsup_{n\to\infty}(1/n)c(x_1^n)\log n$. As an application of Theorem 2.2, we will investigate a property of the encoder based on a distinct parsing of sequences.

Theorem 2.3 Let σ be a length function on $\mathcal{X}$. There exits a sequence $\{\sigma_n^{LZ}\}_{n=1}^{\infty}$ ($\sigma_n^{LZ} : \mathcal{X}^n \to \mathcal{N}$) of length functions satisfying

$$\limsup_{n\to\infty}\frac{\sigma_n^{LZ}(x_1^n)}{n} \le h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x) \quad \forall x \in \mathcal{X}^{\infty}(\sigma).$$

Proof: To prove the theorem, we will construct a variation of the LZ78 code [36]. At first, a given sequence $x \in \mathcal{X}^{\infty}$ is parsed into phrases $w_1w_2w_3\cdots$ inductively according to the following rules:

1. The first word w_1 consists of the single letter x_1.
2. Suppose $w_1\cdots w_j = x_1^{n_j}$.
   (a) If $x_{n_j+1} \notin \{w_1,\cdots,w_j\}$ then w_{j+1} consists of the single letter x_{n_j+1}.

(b) Otherwise, $w_{j+1} = x_{n_j+1}^{m+1}$, where m is the least integer larger than n_j such that $x_{n_j+1}^{m} \in \{w_1, \cdots, w_j\}$ and $x_{n_j+1}^{m+1} \notin \{w_1, \cdots, w_j\}$.

Note that the parsing is sequential, that is, later words have no effect on earlier words. Therefore, an initial segment x_1^n of x can be expressed as

$$x_1^n = w_1 w_2 \cdots w_t v, \tag{2.29}$$

where the final block v is either empty or $v = w_j$ for some $j \le t$. Let $\psi : \mathcal{X} \to \mathcal{B}^*$ be a prefix-free encoder such that

$$\ell(\psi(a)) = \sigma(a) \quad \forall a \in \mathcal{X}.$$

Let $f : \{0, 1, \cdots, n\} \to \mathcal{B}^{\lceil \log n \rceil}$ be a fixed one-to-one function. If x_1^n is parsed as in (2.29), an encoder $\psi_n^{LZ} : \mathcal{X}^n \to \mathcal{B}^*$ maps it into the concatenation

$$\psi_n^{LZ}(x_1^n) = u_1 u_2 \cdots u_t u_{t+1}$$

of binary words $u_j \in \mathcal{B}^*$ ($j = 1, 2, \cdots, t{+}1$), defined according to the following rules:

1. If $j \le t$ and $\ell(w_j) = 1$ then $u_j = 0\psi(w_j)$.
2. If $j \le t$ and $i < j$ is the least integer for which $w_j = w_i a$, $a \in \mathcal{X}$, then $u_j = 1f(i)0\psi(a)$.
3. If v is empty, then u_{t+1} is empty, otherwise $u_{t+1} = 1f(i)$, where i is the least integer such that $v = w_i$.

For each $j = 1, \cdots, t$, let x_{n_j} be the last symbol of w_j. For each j, $0\psi(x_{n_j})$ is appeared in the output binary sequence $\psi_n^{LZ}(x_1^n)$, and requires $1 + \sigma(x_{n_j})$ bits. Moreover, the word $1f(i)$ requires $1 + \lceil \log n \rceil$ bits and appeared in the output binary sequence $\psi_n^{LZ}(x_1^n)$ at most $t + 1$ times. Thus, total codeword length $\ell\left(\psi_n^{LZ}(x_1^n)\right)$ is upper bounded by

$$\begin{aligned} \ell\left(\psi_n^{LZ}(x_1^n)\right) &\le \sum_{j=1}^{t} \left\{1 + \sigma(x_{n_j})\right\} + (t+1)(1 + \lfloor \log n \rfloor) \\ &\le t \log n + \sum_{j=1}^{t} \sigma(x_{n_j}) + 2t + \log n + 2 \end{aligned} \tag{2.30}$$

By the definition of the parsing (2.29), $t \leq c(x_1^n)+1$. Let $B_n = \{n_{j_1}, n_{j_2}, \cdots, n_{j_t}\}$. By Lemma 2.6, we have $c(x_1^n)/n \to 0$, and thus, $|B_n|/n \to 0$. Hence, by combining Lemma 2.7, Lemma 2.4, and (2.30), we have

$$\limsup_{n\to\infty} \frac{1}{n}\ell\left(\psi_n^{LZ}(x_1^n)\right) \leq h(x) \quad \forall x \in \mathcal{X}^\infty(\sigma).$$

Letting $\sigma_n^{LZ}(x_1^n) = \ell\left(\psi_n^{LZ}(x_1^n)\right)$ for each $n \in \mathcal{N}$, we have the theorem. □

As a corollary of Theorem 2.3, we show a sufficient condition for the existence of a sequence $\{\sigma_n^*\}_{n=1}^\infty$ of length functions which is *universal* for a set $A \subset \mathcal{X}^\infty$ in the sense that $\limsup_{n\to\infty}(1/n)\sigma_n^*(x_1^n) \leq \rho^{bl}(x)$ for all $x \in A$.

Corollary 2.1 Let A be a set of sequences ($A \subset \mathcal{X}^\infty$). If there exists a set Σ of countable length functions satisfying $A \subset \bigcup_{\sigma\in\Sigma} \mathcal{X}^\infty(\sigma)$, then there exists a sequence $\{\sigma_n^*\}_{n=1}^\infty$ ($\sigma_n^* : \mathcal{X}^n \to \mathcal{N}$) of length functions such that

$$\limsup_{n\to\infty} \frac{\sigma_n^*(x_1^n)}{n} \leq h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x) \quad \forall x \in A.$$

Proof: We use the technique used in [28] to merge the encoders. Without loss of the generality, we can assume that $\Sigma = \{\sigma^{(1)}, \sigma^{(2)}, \cdots\}$. For each $i = 1, 2, \cdots$, let $\{\sigma_n^{(i)}\}_{n=1}^\infty$ be the sequence of length function constructed from $\sigma^{(i)}$ in Theorem 2.3. Then, we can construct a merged length function $\{\sigma_n^*\}_{n=1}^\infty$ such that

$$\sigma_n^*(x_1^n) = \lceil \log n \rceil + \min_{1\leq i\leq n} \sigma_n^{(i)}(x_1^n).$$

By Theorem 2.3, we have

$$\limsup_{n\to\infty} \frac{\sigma_n^*(x_1^n)}{n} \leq h(x) \quad \forall x \in \bigcup_{i=1}^\infty \mathcal{X}^\infty(\sigma^{(i)}).$$

Since $A \subset \bigcup_{\sigma\in\Sigma} \mathcal{X}^\infty(\sigma) = \bigcup_{i=1}^\infty \mathcal{X}^\infty(\sigma^{(i)})$, we have the corollary. □

In the following, we investigate a relation between the condition shown by Corollary 2.1 and the Kieffer's condition for the existence of a universal encoder for a set of stationary ergodic sources.

Let Ω be a set of stationary ergodic sources. A sequence $\{\sigma_n\}_{n=1}^\infty$ ($\sigma_n : \mathcal{X}^n \to \mathcal{N}$) of length functions is said to be *weakly universal* for Ω, if

$$\limsup_{n\to\infty} E\left[\sigma_n(X_1^n)\right] = H(\boldsymbol{X}) \quad \forall \boldsymbol{X} = \{X_i\}_{i=1}^\infty \in \Omega.$$

Kieffer [10] showed the necessary and sufficient condition for the existence of sequence of length functions which is weakly universal for a set Ω of stationary ergodic sources.

Theorem 2.4 (Kieffer's condition [10]) Let Ω be a set of stationary ergodic sources. There exists a sequence $\{\sigma_n\}_{n=1}^{\infty}$ of length functions which is weakly universal for Ω, if and only if there exits a set $\mathcal{P}$ of countable probability distributions on $\mathcal{X}$ such that for any $\boldsymbol{X} \in \Omega$ with $H(\boldsymbol{X}) < \infty$, there exists $p \in \mathcal{P}$ satisfying

$$E\left[-\log p(X_1)\right] < \infty.$$

Suppose that there exists a universal encoder for a set Ω of stationary ergodic sources. Then, the next theorem shows that there exists a universal encoder for a set of individual sequences defined by Ω.

Theorem 2.5 Let Ω be a set of stationary ergodic sources. Suppose there exists a universal source code σ^* for Ω. Then there exist a set $A^* \subset \mathcal{X}^{\infty}$ and a sequence of length functions $\{\sigma_n^*\}_{n=1}^{\infty}$ ($\sigma_n^* : \mathcal{X}^n \to \mathcal{N}$) satisfying that

$$P_{X_1^{\infty}}(A^*) = 1,$$

for any $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty} \in \Omega$ with $H(\boldsymbol{X}) < \infty$, and that

$$\limsup_{n\to\infty} \frac{\sigma_n^*(x_1^n)}{n} \leq h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x) \quad \forall x \in A^*.$$

Before proving the theorem, we state a key lemma.

Lemma 2.12 Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source. If there exists a length function σ on $\mathcal{X}$ satisfying

$$E[\sigma(X_1)] < \infty, \tag{2.31}$$

then, for almost all $x \in \mathcal{X}^{\infty}$, we have

$$\lim_{u\to\infty} \lim_{n\to\infty} \sum_{a:\sigma(a)\geq u} p_1(a|x_1^n)\sigma(a) = 0.$$

Proof: To prove the lemma, we show that there exists a set $A^{(\sigma)} \subset \mathcal{X}^{\infty}$ satisfying the following two properties: (i) $P_{X_1^{\infty}}(A^{(\sigma)}) = 1$ and (ii) for any

$x \in A^{(\sigma)}$ and any $\gamma > 0$, there exists $u_0 = u_0(x, \gamma)$ and $n_0 = n_0(x, \gamma)$ such that

$$\sum_{a:\sigma(a) \geq u} p_1(a|x_1^n)\sigma(a) < \gamma \quad \forall u \geq u_0 \text{ and } \forall n \geq n_0.$$

Fix $\gamma > 0$. Since the condition (2.31) implies

$$\lim_{u \to \infty} \sum_{a:\sigma(a) \geq u} P_{X_1}(a)\sigma(a) = 0,$$

we can choose $u_1 = u_1(\boldsymbol{X}, \gamma)$ so that

$$\sum_{a:\sigma(a) \geq u} P_{X_1}(a)\sigma(a) < \gamma/2 \quad \forall u \geq u_1.$$

Define the function f_u on $\mathcal{X}$ by

$$f_u(a) \triangleq \begin{cases} \sigma(a) & : \text{ if } \sigma(a) \geq u, \\ 0 & : \text{ otherwise.} \end{cases}$$

By the ergodic theorem [5], we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f_u(x_i) = E[f_u(X_1)] \quad \text{a.s.}$$

that is,

$$\lim_{n \to \infty} \sum_{a:\sigma(a) \geq u} p_1(a|x_1^n)\sigma(a) = \sum_{a:\sigma(a) \geq u} P_{X_1}(a)\sigma(a) \quad \text{a.s.}$$

Hence, for almost all $x \in \mathcal{X}^\infty$, we can choose $n_1 = n_1(x, \gamma)$ so that

$$\sum_{a:\sigma(a) \geq u} p_1(a|x_1^n)\sigma(a) < \gamma \quad \forall u \geq u_1 \text{ and } \forall n \geq n_1.$$

This implies that for any $\gamma > 0$, there exists $A_\gamma \subset \mathcal{X}^\infty$ satisfying $P_{X_1^\infty}(A_\gamma) = 1$ and

$$\lim_{u \to \infty} \lim_{n \to \infty} \sum_{a:\sigma(a) \geq u} p_1(a|x_1^n)\sigma(a) < \gamma \quad \forall x \in A_\gamma.$$

Let $A^{(\sigma)} = \cap_{m=1}^{\infty} A_{1/m}$. Then, $A_{1/1} \supset A_{1/2} \supset \cdots$ and thus,

$$P_{X_1^\infty}(A^{(\sigma)}) = P_{X_1^\infty}(\cap_{m=1}^{\infty} A_{1/m}) = \lim_{m \to \infty} P_{X_1^\infty}(A_{1/m}) = 1.$$

Note that for any $\gamma > 0$, there exists $m \in \mathcal{N}$ such that $1/m < \gamma$. Hence, for all $x \in A^{(\sigma)}$ and any $\gamma > 0$, there exists $u_0 = u_0(x, \gamma) = u_1(\boldsymbol{X}, 1/m)$ and $n_0 = n_0(x, \gamma) = n_1(x, 1/m)$ such that

$$\sum_{a:\sigma(a) \geq u} p_1(a|x_1^n)\sigma(a) < 1/m < \gamma \quad \forall u \geq n_0 \text{ and } \forall n \geq n_0.$$

Hence, we have the lemma. $\square$

Proof of Theorem 2.5: Since there exists a sequence of length functions which is universal for Ω, there exits a set $\mathcal{P}$ of countable probability distributions on $\mathcal{X}$ such that for any $\boldsymbol{X} \in \Omega$ with $H(\boldsymbol{X}) < \infty$, there exists $p \in \mathcal{P}$ satisfying

$$E\left[-\log p(X_1)\right] < \infty.$$

For each $p \in \mathcal{P}$, we can construct a length function σ_p on $\mathcal{X}$ such that

$$-\log p(a) + 1 \leq \sigma_p(a) < -\log p(a) + 2 \quad \forall a \in \{a' : p(a') > 0\}.$$

The above discussion implies that for each $\boldsymbol{X} \in \Omega$ with $H(\boldsymbol{X}) < \infty$, there exists a $p \in \mathcal{P}$ such that $E\left[-\log p(X_1)\right] < \infty$, and thus,

$$E\left[\sigma_p(X_1)\right] \leq E\left[-\log p(X_1) + 2\right] < \infty.$$

Hence, by Lemma 2.12, there exits a set $A^{(\sigma_p)}$ such that $P_{X_1^\infty}(A^{(\sigma_p)}) = 1$ and $A^{(\sigma_p)} \subset \mathcal{X}^\infty(\sigma_p)$. By Corollary 2.1, we can show that $A^* \triangleq \bigcup_{p\in\mathcal{P}} A^{(\sigma_p)}$ satisfies the theorem. $\Box$

Now, we give a corollary of Theorem 2.5. Let $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ be a stationary ergodic source such that $H(\boldsymbol{X}) < \infty$. By the ergodic theorem, for almost all $x \in \mathcal{X}^\infty$,

$$\hat{H}(x) = H(\boldsymbol{X}).$$

Further, assume that $\boldsymbol{X}$ satisfies the assumption of Lemma 2.12. Then, by Lemma 2.12, we have

$$h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x) = H(\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$. The above discussion gives the following corollary.

Corollary 2.2 Let Ω be a set of stationary ergodic sources appeared in Theorem 2.5. Then, for any $\boldsymbol{X} \in \Omega$ with $H(\boldsymbol{X}) < \infty$, we have

$$h(x) = \hat{H}(x) = \tilde{H}(x) = \rho^{bl}(x) = H(\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$.

Chapter 3

Fixed-Rate Lossy Coding Problem

3.1 Introduction

In Chapter 2, we considered the lossless coding problem of individual sequences over countably infinite source alphabet, and investigated the optimal performance theoretically attainable by blockwise lossless coding schemes. However, when the source alphabet is not finite or countable, the framework of lossless coding is not valid. For example, the description of an arbitrary real number requires an infinite number of bits, so a finite representation of a continuous random variable cannot be perfect. As an alternative approach, we need to permit finite distortion between a input data and the reproduced data within some acceptable level. Clearly, a smaller distortion can be achieved if we are allowed to use a higher coding rate. In this and the following chapters (Chapter 3–5), we investigate the asymptotic optimal trade-off between the distortion and the coding rate for a given individual sequence attainable by blockwise coding schemes.

Historically, the lossy coding problem of individual sequences was formulated by Ziv [34]. He studied the fixed-rate lossy coding problem of individual sequences with finite alphabet, and presented a universal lossy coding at fixed-rate. Following the same line as in [34], Yang and Kieffer [31] studied the fixed-rate and fixed-distortion lossy coding problems, and proposed two universal lossy coding schemes based on LZ78 code. Their result shows that the optimal distortion attainable by fixed-rate encoders is equal to the smallest topological entropy among reproduction sequences satisfying the distortion criterion, and that the optimal rate attainable by fixed-distortion encoders is equal to the optimal average rate with respect to the non-overlapping empirical distribution of the given sequence.

In this chapter, we investigate the fixed-rate lossy coding problem of individual sequences subject to the subadditive distortion measure. We formulate the fixed-rate lossy coding problem of individual sequences as the problem of optimizing the distortion over all the blockwise fixed-rate lossy encoders. Then, we characterize the optimal performance theoretically achievable (OPTA) function for the given individual sequence by using the empirical distribution of the sequence. Our result clarifies that the optimal distortion achievable by fixed-rate lossy coding schemes is equal to the optimal average distortion with respect to the overlapping empirical distribution of the given sequence. It is also proved that, instead of overlapping empirical distribution, we can adopt the non-overlapping empirical distribution to characterize the OPTA function. Corollary of main result reveals that the optimal distortion is equal to the distortion-rate function derived by Yang and Kieffer. To prove the coding theorem, we propose the fixed-rate lossy coding scheme based on the complexity of the sequence, and show that the scheme is universal, that is, achieves OPTA for any sequence under some mild conditions. This result reveals that not only LZ78 code, but various universal lossless codes such as grammar based codes can also be applied to universal lossy coding.

As an application of main result, in Section 3.5, we investigate the fixed-rate lossy coding problem of stationary ergodic sources, and reveal that the blockwise coding schemes can attain the optimal distortion for the source.

3.2 Fixed-Rate Lossy Coding of Individual Sequences

In this chapter, unless stated otherwise, the source alphabet $\mathcal{X}$ is assumed to be arbitrary set. We introduce another set $\mathcal{Y}$ called the *reproduction alphabet*. The size of $\mathcal{Y}$ is assumed to be finite throughout this chapter.

For each $n \in \mathcal{N}$, $d_n(x_1^n, y_1^n)$ is called the *distortion* between $x_1^n \in \mathcal{X}^n$ and $y_1^n \in \mathcal{Y}^n$, where d_n is any mapping from $\mathcal{X}^n \times \mathcal{Y}^n$ into $[0, +\infty)$ and is called the *distortion measure*. The distortion per symbol is given by

$$\frac{1}{n} d_n(x_1^n, y_1^n) \geq 0.$$

In this chapter, unless otherwise specified, we assume that the distortion measure d_n is *subadditive*, that is, for any $m \in \mathcal{N}$, $n \in \mathcal{N}$, $x_1^{m+n} \in \mathcal{X}^{m+n}$, and $y_1^{m+n} \in \mathcal{Y}^{m+n}$,

$$d_{m+n}(x_1^{m+n}, y_1^{m+n}) \leq d_m(x_1^m, y_1^m) + d_n(x_{m+1}^{m+n}, y_{m+1}^{m+n}). \tag{3.1}$$

A fixed-rate lossy encoder of block size k is defined by a *quantizer* $\pi_k : \mathcal{X}^k \to \mathcal{Y}^k$. By a fixed-rate lossy block encoder, a given sequence $x \in \mathcal{X}^\infty$ is encoded in the following manner. At first, x is parsed into blocks $x_{tk+1}^{(t+1)k}$ $(t = 0, 1, \cdots)$ of size k. Next, each block $x_{tk+1}^{(t+1)k}$ is quantized into $y_{tk+1}^{(t+1)k} = \pi_k(x_{tk+1}^{(t+1)k})$ by the quantizer π_k. Then, each $y_{tk+1}^{(t+1)k}$ is represented with $\lceil \log \|\pi_k\| \rceil$ bits, where $\|\pi_k\| \triangleq |\{\pi_k(a_1^k) : a_1^k \in \mathcal{X}^k\}|$. The rate $\mathrm{rate}(\pi_k)$ of the quantizer π_k is defined as

$$\mathrm{rate}(\pi_k) \triangleq \frac{1}{k} \lceil \log \|\pi_k\| \rceil.$$

For each sequence $x \in \mathcal{X}^\infty$, the asymptotic distortion attained by the quantizer π_k is defined as

$$\delta(x, \pi_k) \triangleq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{k} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right).$$

Then, the optimal distortion $\delta(R|x)$ attainable by blockwise lossy coding at the fixed-rate $R > 0$ is defined as

$$\delta(R|x) \triangleq \inf_k \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k) \leq R}} \delta(x, \pi_k),$$

where $\inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k) \leq R}}$ is taken over all the π_k satisfying $\mathrm{rate}(\pi_k) \leq R$.

In order to clarify the optimal distortion $\delta(R|x)$, we introduce the distortion-rate function. Let $p_k(\cdot|x_1^n)$ be the overlapping empirical distribution of x_1^n defined in (2.1). Then, $\hat{D}_k(R|x_1^n)$ is defined as the optimal average distortion with respect to $p_k(\cdot|x_1^n)$:

$$\begin{aligned} \hat{D}_k(R|x_1^n) &\triangleq \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k) \leq R}} E_p\left[\frac{1}{k} d_k\left(X_1^k, \pi_k(X_1^k)\right)\right] \\ &= \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k) \leq R}} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \frac{1}{k} d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right), \end{aligned}$$

where inf is taken over the all π_k satisfying $\mathrm{rate}(\pi_k) \leq R$. Lastly, the *distortion-rate function* $\hat{D}(R|x)$ of the sequence $x \in \mathcal{X}^\infty$ is defined as

$$\hat{D}(R|x) \stackrel{\triangle}{=} \liminf_{k\to\infty} \limsup_{n\to\infty} \hat{D}_k(R|x_1^n).$$

The following theorem gives a lower bound on the distortion attainable by the fixed-rate block encoders.

Theorem 3.1 (Converse coding theorem) Assume that for a given sequence $x \in \mathcal{X}^\infty$, there exists a *reference letter* $r \in \mathcal{Y}$ satisfying that

$$\lim_{d\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{\substack{1\leq i\leq n \\ d_1(x_i,r)\geq d}} d_1(x_i, r) = 0. \tag{3.2}$$

Then, for any $R > 0$ and any block encoder π_k of the block size k satisfying $\mathrm{rate}(\pi_k) < R$, we have

$$\delta(x, \pi_k) \geq \hat{D}(R|x). \tag{3.3}$$

Remark 3.1 The assumption of existence of the reference letter $r \in \mathcal{Y}$ satisfying (3.2) corresponds to the commonly required assumption in a theory of lossy coding of probabilistic sources, that is, for a source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$, there exists a reference letter $r \in \mathcal{Y}$ satisfying $E[d(X_1, r)] < \infty$. As shown in Lemma 3.5, if $\boldsymbol{X}$ is a stationary ergodic source for which $E[d(X_1, r)] < \infty$ holds, then r satisfies (3.2) for almost all $x \in \mathcal{X}^\infty$.

Next, we show the coding theorem by constructing a universal fixed-rate lossy encoder. To define the universal lossy block encoder, we introduce the complexity function.

Definition 3.1 (Complexity function) A *complexity function* $L : \mathcal{Y}^* \to \mathcal{N}$ is a function which satisfies the following two properties:

1. For any $n \in \mathcal{N}$,
$$\sum_{y_1^n \in \mathcal{Y}^n} 2^{-L(y_1^n)} \leq 1. \tag{3.4}$$

2. For any K and any $\gamma > 0$, there exists an integer $n_0 = n_0(\gamma, K, \mathcal{Y}, L)$ such that for any $n \geq n_0$ and any finite-state IL encoder ψ with state set $\mathcal{S}$ of K states,
$$L(y_1^n) \leq \ell\left(\psi(y_1^n)\right) + n\gamma \quad \forall y_1^n \in \mathcal{Y}^n, \tag{3.5}$$

where $\psi(y_1^n) \equiv f(s_1, y_1) f(s_2, y_2) \cdots f(s_n, y_n)$ denotes the output binary sequence when y_1^n is encoded by ψ.

Remark 3.2 Muramatsu and Kanaya [20] defined the complexity function as a function satisfying the condition 1 in Definition 3.1 and

$$\limsup_{n\to\infty} \frac{1}{n} L(Y_1^n) \leq H(\boldsymbol{Y}) \quad \text{a.s.},$$

for any stationary ergodic source $\boldsymbol{Y} = \{Y_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{Y}$. However, in this thesis, we adopt an alternative definition in which stationary ergodic sources are not involved. This allows us to deal with any individual sequence which may not be an output sequence from an stationary ergodic source.

Example 3.1 The length function of the LZ78 code [36] satisfies both conditions in Definition 3.1, and thus is a complexity function. Moreover, the length function of any grammar based code G [14] also satisfies (3.5) [29], and is a complexity function.

Let L be an arbitrary complexity function. For each $k \in \mathcal{N}$ and $R > 0$, let π_k^R be the quantizer defined as

$$\pi_k^R(a_1^k) \triangleq \arg \min_{\substack{b_1^k \in \mathcal{Y}^k \\ L(b_1^k) \leq kR}} d_k(a_1^k, b_1^k) \quad \forall a_1^k \in \mathcal{X}^k. \tag{3.6}$$

By the definition of π_k^R and the property (3.4) of the complexity function, it is clear that $\mathrm{rate}(\pi_k^R) \leq R$. The next theorem shows that the quantizer π_k^R is optimum.

Theorem 3.2 (Coding theorem) For any $x \in \mathcal{X}^{\infty}$ and any $R > 0$, the sequence $\{\pi_k^R\}_{k=1}^{\infty}$ of quantizers satisfies that

$$\limsup_{k\to\infty} \delta(x, \pi_k^R) \leq \lim_{\zeta \downarrow 0} \hat{D}(R - \zeta | x), \tag{3.7}$$

provided that there exists a reference letter r satisfying (3.2).

Theorem 3.1 and Theorem 3.2 will be proved in Section 3.4. According to Theorem 3.1 and Theorem 3.2, we have the following corollary.

Corollary 3.1 For any $x \in \mathcal{X}^\infty$, we have

$$\lim_{k\to\infty} \delta(x, \pi_k^R) = \delta(R|x) = \hat{D}(R|x),$$

provided that $\hat{D}(R|x)$ is continuous at the point R, and there exists a reference letter r satisfying (3.2).

3.3 Relations among Distortion-Rate Functions of Individual Sequence

The distortion-rate function may be defined by using the non-overlapping empirical distribution $q_k(\cdot|x_1^n)$ defined in (2.3) instead of the overlapping empirical distribution $p_k(\cdot|x_1^n)$. Let $\tilde{D}_k(R|x_1^{Tk})$ be the optimal average distortion with respect to $q_k(\cdot|x_1^{Tk})$:

$$\begin{aligned}\tilde{D}_k(R|x_1^{Tk}) &\triangleq \inf_{\substack{\pi_k\\ \text{rate}(\pi_k)\le R}} E_q\left[\frac{1}{k} d_k\left(X_1^k, \pi_k(X_1^k)\right)\right] \\ &= \inf_{\substack{\pi_k\\ \text{rate}(\pi_k)\le R}} \frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{k} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right).\end{aligned}$$

Then, the non-overlapping version $\tilde{D}(R|x)$ of the distortion-rate function of the sequence $x \in \mathcal{X}^\infty$ is defined as

$$\tilde{D}(R|x) \triangleq \liminf_{k\to\infty} \limsup_{T\to\infty} \tilde{D}_k(R|x_1^{Tk}).$$

The next theorem corresponds to Corollary 3.1.

Theorem 3.3 For any $x \in \mathcal{X}^\infty$, we have

$$\lim_{k\to\infty} \delta(x, \pi_k^R) = \delta(R|x) = \tilde{D}(R|x),$$

provided that $\tilde{D}(R|x)$ is continuous at the point R, and there exists a reference letter r satisfying (3.2).

Theorem 3.3 will be proved in Section 3.4.

Yang and Kieffer [31] investigated the fixed-rate lossy block coding of individual sequences under the assumption that the distortion measure is *additive*, that is, for any $m \in \mathcal{N}$, $n \in \mathcal{N}$, $x_1^{m+n} \in \mathcal{X}^{m+n}$, and $y_1^{m+n} \in \mathcal{Y}^{m+n}$,

$$d_{m+n}(x_1^{m+n}, y_1^{m+n}) = d_m(x_1^m, y_1^m) + d_n(x_{m+1}^{m+n}, y_{m+1}^{m+n}).$$

They defined the distortion-rate function $D^{YK}(R|x)$ as the infimum of D such that there exists $y \in \mathcal{Y}^\infty$ satisfying the distortion constraint

$$\limsup_{n\to\infty} \frac{1}{n} d_n(x_1^n, y_1^n) \leq D,$$

and the rate constraint

$$h(y) \leq R,$$

where $h(y)$ is the *topological entropy* of y [39] defined by

$$h(y) \triangleq \lim_{m\to\infty} \frac{1}{m} \log |B_m(y)|,$$

and

$$B_m(y) \triangleq \left\{ b_1^m \in \mathcal{Y}^m : \exists i \text{ s.t. } y_i^{i+m-1} = b_1^m \right\}.$$

Moreover, Yang and Kieffer [31] proposed the universal quantizer π_k^{LZ} based on the length function L^{LZ} of the LZ78 code, and showed

$$\lim_{k\to\infty} \delta(x, \pi_k^{LZ}) = D^{YK}(R|x),$$

provided that $D^{YK}(\cdot|x)$ is continuous at the point R.

Since L^{LZ} is a complexity function, Yang and Kieffer's quantizer π_k^{LZ} is a special case of our quantizer π_k^R. Combining their result and Corollary 3.1, we have the following corollary. The following corollary shows that the distortion-rate function $\hat{D}(R|x)$ is equal to the distortion-rate function $D^{YK}(R|x)$ defined by Yang and Kieffer.

Corollary 3.2 Assume that the distortion measure d_n is additive. For any $x \in \mathcal{X}^\infty$, we have

$$\hat{D}(R|x) = D^{YK}(R|x),$$

provided that both of $\hat{D}(\cdot|x)$ and $D^{YK}(\cdot|x)$ are continuous at the point R, and there exists a reference letter r satisfying (3.2).

3.4 Proofs of Theorems

In this section, we shall prove Theorem 3.1, Theorem 3.2, and Theorem 3.3. Before proving the theorems, we introduce some lemmas.

At first, we show an important property of the complexity function. The following lemma shows that any complexity function can be characterized as a lower bound on the codeword length of the blockwise lossless encoder.

Lemma 3.1 Let L be a complexity function. For any $m \in \mathcal{N}$, any length function σ_m on $\mathcal{Y}^m$, and any $\epsilon > 0$, there exists an integer $n_0 = n_0(\epsilon, m, \mathcal{Y}, L)$ such that for any $n \geq n_0$ and $y_1^n \in \mathcal{Y}^n$,

$$L(y_1^n) \leq \min_{0 \leq j \leq m-1} \sum_{\substack{0 \leq i \leq n-m \\ i \equiv j \bmod m}} \sigma_m(y_{i+1}^{i+m}) + n\epsilon.$$

Proof: This lemma can be proved by a similar manner as the proof of [31, Lemma 1].

For each j, let τ_j be the length function such that $\tau_j(b_1^j) = \lceil j \log |\mathcal{Y}| \rceil$ for any $b_1^j \in \mathcal{Y}^j$. We can construct the finite-state IL encoder ψ_j which encode a given sequence $y \in \mathcal{Y}^\infty$ in the following manner: y is parsed into blocks of length $j, m, m, \cdots$, and the first block is encoded with $\tau_j(y_1^j)$ bits and remaining blocks y_{i+1}^{i+m} $(i = j, j+m, j+2m, \cdots)$ is encoded with $\sigma_m(y_{i+1}^{i+m})$ bits. Hence, we have

$$\ell(\psi_j(y_1^n)) \leq \tau_j(y_1^j) + \sum_{\substack{0 \leq i \leq n-m+1 \\ i \equiv j \bmod m}} \sigma_m(y_{i+1}^{i+m}). \tag{3.8}$$

Fix $\gamma > 0$. The condition (3.5) of the complexity function and (3.8) implies that there exits $n_1 \in \mathcal{N}$ such that for any $n \geq n_1$ and $y_1^n \in \mathcal{Y}^n$,

$$\begin{aligned} L(y_1^n) &\leq \tau_j(y_1^j) + \sum_{\substack{0 \leq i \leq n-m \\ i \equiv j \bmod m}} \sigma_m(y_{i+1}^{i+m}) + M + n\gamma \\ &= \sum_{\substack{0 \leq i \leq n-m \\ i \equiv j \bmod m}} \sigma_m(y_{i+1}^{i+m}) + n\left(\frac{\tau_j(y_1^j)}{n} + \frac{M}{n} + \gamma\right), \end{aligned}$$

where $M = \max_{b_1^m \in \mathcal{Y}^m} \sigma_m(b_1^m)$. Hence, we have the lemma. $\square$

Next, we show some properties of reference letter.

Lemma 3.2 For any $x \in \mathcal{X}^\infty$ and $r \in \mathcal{Y}$ satisfying (3.2), we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} d_1(x_i, r) < \infty.$$

Proof: Since (3.2) holds, for any $\gamma > 0$ and sufficiently large d and n, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} d_1(x_i, r) &\leq \frac{1}{n}\sum_{i=1}^{n} d + \frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ d_1(x_i, r) \geq d}} d_1(x_i, r) \\ &\leq d + \gamma, \end{aligned}$$

and thus,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} d_1(x_i, r) < \infty.$$

Hence, we have the lemma. □

Lemma 3.3 Assume that $x \in \mathcal{X}^\infty$ and $r \in \mathcal{Y}$ satisfies (3.2). Then, for any $\gamma > 0$, there exists $\epsilon > 0$ and $n_0 \in \mathcal{N}$ such that for any $n \geq n_0$ and any set $B \subset \{1, \cdots, n\}$ satisfying $|B|/n < \epsilon$,

$$\frac{1}{n}\sum_{i\in B} d_1(x_i, r) \leq \gamma.$$

Proof: Since (3.2) holds, for any $\gamma' > 0$ and sufficiently large d and n, we have

$$\frac{1}{n}\sum_{\substack{1\leq i\leq n\\ d_1(x_i,r)\geq d}} d_1(x_i, r) \leq \gamma'.$$

Choose $\gamma' > 0$ and $\epsilon > 0$ so that $\epsilon d + \gamma' < \gamma$. Then, we have,

$$\begin{aligned}
\frac{1}{n}\sum_{i\in B} d_1(x_i, r) &\leq \frac{1}{n}\sum_{\substack{i\in B\\ d_1(x_i,r)\leq d}} d_1(x_i, r) + \frac{1}{n}\sum_{\substack{i\in B\\ d_1(x_i,r)\geq d}} d_1(x_i, r)\\
&\leq \frac{1}{n}\sum_{i\in B} d + \frac{1}{n}\sum_{d_1(x_i,r)\geq d} d_1(x_i, r)\\
&\leq \epsilon d + \gamma'\\
&\leq \gamma,
\end{aligned}$$

and thus, the lemma is proved. □

Now, we show a lemma which plays an important role in the proofs. For each $m \in \mathcal{N}$, $k \in \mathcal{N}$, $a_1^m \in \mathcal{X}^m$, and $r \in \mathcal{Y}$, define $\Delta_k(a_1^m, r)$ as

$$\Delta_k(a_1^m, r) \triangleq \sum_{i=1}^{k-1} d_1(a_i, r) + \sum_{i=m-k+2}^{m} d_1(a_i, r). \tag{3.9}$$

Then, we have the following lemma.

Lemma 3.4 Let L be a complexity function and fix $r \in \mathcal{Y}$. For any $R > 0$, $\epsilon > 0$, and any integer k satisfying $k > 1/\epsilon$, there exists $m_0 = m_0(\epsilon, k, \mathcal{Y}, L)$ such that for any π_k satisfying rate$(\pi_k) \leq R - \epsilon$ and any $m \geq m_0$, the quantizer π_m^R defined in (3.6) satisfies that

$$d_m\left(a_1^m, \pi_m^R(a_1^m)\right) \leq \min_{0\leq j\leq k-1} \sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} d_k\left(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})\right) + \Delta_k(a_1^m, r) \quad \forall a_1^m \in \mathcal{X}^m,$$

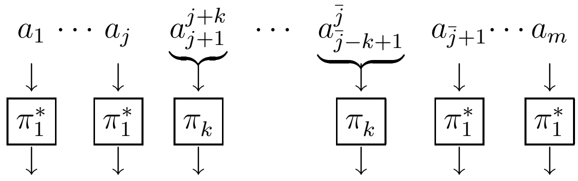

Figure 3.1: Construction of $\pi_{k,m}^{(j)}$

and thus,

$$d_m\left(a_1^m, \pi_m^R(a_1^m)\right) \le \sum_{i=0}^{m-k} \frac{1}{k} d_k\left(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})\right) + \Delta_k(a_1^m, r) \quad \forall a_1^m \in \mathcal{X}^m.$$

Proof: For a given quantizer π_k and each j ($0 \le j \le k-1$), let $\pi_{k,m}^{(j)}$ be the quantizer which quantizes a block $a_1^m \in \mathcal{X}^m$ by the following manner. At first, the first j symbols $a_1, \cdots, a_j$ of a_1^m are quantized by the quantizer π_1^* such that $\pi_1^*(a) = r$ for all $a \in \mathcal{X}$. Then, the following a_{j+1}^m is partitioned into blocks of size k and each block is quantized by π_k. Lastly, the remaining symbols $a_{\bar{j}+1}, \cdots, a_m$ ($m - \bar{j} < k$) of a_1^m are quantized by π_1^* (see Fig. 3.1). Since d_k is subadditive, it is apparent that for any π_k and j, $\pi_{k,m}^{(j)}$ satisfies that

$$\begin{aligned} d_m\left(a_1^m, \pi_{k,m}^{(j)}(a_1^m)\right) &\le \sum_{i=1}^{j} d_1(a_i, \pi_1^*(a_i)) \\ &\quad + \sum_{\substack{i \equiv j \bmod k \\ 0 \le i \le m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) \\ &\quad + \sum_{i=\bar{j}+1}^{m} d_1(a_i, \pi_1^*(a_i)) \\ &\le \sum_{\substack{i \equiv j \bmod k \\ 0 \le i \le m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) + \Delta_k(a_1^m, r) \end{aligned} \tag{3.10}$$

Fix γ so that $0 < \gamma \le \epsilon - 1/k$. Let $\bar{\sigma}_k$ be a length function on $\mathcal{Y}^k$ defined as

$$\bar{\sigma}_k(b_1^k) = \begin{cases} \lceil \log \|\pi_k\| \rceil + 1 & : \text{if } b_1^k \in \{\pi_k(a_1^k) : a_1^k \in \mathcal{X}^k\} \\ \lceil \log |\mathcal{Y}^k| \rceil + 1 & : \text{otherwise.} \end{cases}$$

Then, by Lemma 3.1, there exists $m_1 = m_1(\gamma, k, \mathcal{Y}, L)$ such that for any

$m \geq m_1$ and any $j = 0, \cdots, k-1$,

$$L(b_1^m) \leq \sum_{\substack{0 \leq i \leq m-k \\ i \equiv j \bmod k}} \bar{\sigma}_k(b_{i+1}^{i+k}) + m\gamma \quad \forall b_1^m \in \mathcal{Y}^m.$$

Thus, if $m \geq m_1$, for any $a_1^m \in \mathcal{X}^m$,

$$\begin{aligned} L\left(\pi_{k,m}^{(j)}(a_1^m)\right) &\leq \frac{m}{k}(\lceil \log \|\pi_k\| \rceil + 1) + m\gamma \\ &\leq mR + m\left(\frac{1}{k} - \epsilon + \gamma\right) \\ &\leq mR. \end{aligned}$$

Hence, for all $j = 0, \cdots, k-1$, we have

$$L\left(\pi_{k,m}^{(j)}(a_1^m)\right) \leq mR \quad \forall a_1^m \in \mathcal{X}^m. \tag{3.11}$$

According to (3.11), the definition of π_k^R and (3.10) yields

$$\begin{aligned} d_m\left(a_1^m, \pi_m^R(a_1^m)\right) &\leq \min_{0 \leq j \leq k-1} d_m\left(a_1^m, \pi_{k,m}^{(j)}(a_1^m)\right) \\ &\leq \min_{0 \leq j \leq k-1} \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} d_k\left(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})\right) + \Delta_k(a_1^m, r). \end{aligned}$$

This completes the proof of the lemma. □

Proof of Theorem 3.1: Notice that

$$\begin{aligned} &\hat{D}(R|x) \\ &= \liminf_{m \to \infty} \limsup_{n \to \infty} \inf_{\substack{\pi_m \\ \text{rate}(\pi_m) \leq R}} \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \frac{1}{m} d_m\left(x_i^{i+m-1}, \pi_m(x_i^{i+m-1})\right) \\ &\leq \limsup_{m \to \infty} \limsup_{n \to \infty} \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \frac{1}{m} d_m\left(x_i^{i+m-1}, \pi_m^R(x_i^{i+m-1})\right). \end{aligned}$$

Hence, to show the theorem, it is sufficient to show that

$$\begin{aligned} &\delta(x, \pi_{k'}) \\ &\geq \limsup_{m \to \infty} \limsup_{n \to \infty} \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \frac{1}{m} d_m\left(x_i^{i+m-1}, \pi_m^R(x_i^{i+m-1})\right), \end{aligned} \tag{3.12}$$

for any quantizer $\pi_{k'}$ such that $\text{rate}(\pi_{k'}) < R$.

Fix a quantizer $\pi_{k'}$ and $\epsilon > 0$ such that $\mathrm{rate}(\pi_{k'}) \le R-\epsilon < R$. Let $w \in \mathcal{N}$ be sufficiently large so that $k \equiv wk'$ satisfies $k > 1/\epsilon$. We can construct the quantizer π_k of block size $k = wk'$ which quantizes a block $a_1^k \in \mathcal{X}^k$ by the following manner. At first, a_1^k is partitioned into w blocks of size k', and then, each block is quantized by $\pi_{k'}$. By the definition, $\mathrm{rate}(\pi_k) \le \mathrm{rate}(\pi_{k'}) \le R-\epsilon$ and $\delta(x, \pi_k) \le \delta(x, \pi_{k'})$.

Fix $\gamma > 0$, and choose T_0 such that for any $T \ge T_0$,

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) \le \delta(x, \pi_k) + \gamma. \tag{3.13}$$

By Lemma 3.4, there exists $m_0 = m_0(\epsilon, k, \mathcal{Y}, L)$ such that for any $m \ge m_0$, we have

$$\begin{aligned} & d_m\left(x_i^{i+m-1}, \pi_m^R(x_i^{i+m-1})\right) \\ \le\ & \min_{0\le j\le k-1} \sum_{\substack{l\equiv j \bmod k \\ 0\le l\le m-k}} d_k\left(x_{l+i}^{l+k+i-1}, \pi_k(x_{l+i}^{l+k+i-1})\right) + \Delta_k(x_i^{i+m-1}, r) \end{aligned}$$

for any $i \in \mathcal{N}$. Especially, we have

$$\begin{aligned} & d_m\left(x_i^{i+m-1}, \pi_m^R(x_i^{i+m-1})\right) \\ \le\ & \sum_{\substack{l\equiv j_i \bmod k \\ 0\le l\le m-k}} d_k\left(x_{l+i}^{l+k+i-1}, \pi_k(x_{l+i}^{l+k+i-1})\right) + \Delta_k(x_i^{i+m-1}, r) \\ =\ & \sum_{\substack{t \\ i\le (tk+1)\le i+m-k}} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \Delta_k(x_i^{i+m-1}, r) \end{aligned}$$

where j_i is chosen so that $i + j_i \equiv 1 \bmod k$. Hence, we have

$$\begin{aligned} & \sum_{i=1}^{n-m+1}\frac{1}{m}d_m\left(x_i^{i+m-1}, \pi_m^R(x_i^{i+m-1})\right) \\ \le\ & \frac{1}{m}\sum_{i=1}^{n-m+1}\sum_{\substack{t \\ i\le (tk+1)\le i+m-k}} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \frac{1}{m}\sum_{i=1}^{n-m+1}\Delta_k(x_i^{i+m-1}, r) \\ \le\ & \sum_{t=0}^{\lfloor n/k\rfloor-1} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \frac{k}{m}\sum_{i=1}^{n} d_1(x_i, r), \end{aligned} \tag{3.14}$$

where the last inequality follows from that each k-blocks is encoded by π_k at most m times. Since (3.14) holds for any n and T ($n \ge T$), letting $T = \lfloor n/k \rfloor$, we have

$$\begin{aligned}
&\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}d_m\left(x_i^{i+m-1},\pi_m^R(x_i^{i+m-1})\right)\\
&\leq \quad \frac{1}{n-m+1}\sum_{t=0}^{T-1}d_k\left(x_{tk+1}^{(t+1)k},\pi_k(x_{tk+1}^{(t+1)k})\right)\\
&\quad +\frac{k}{m(n-m+1)}\sum_{i=1}^{n}d_1(x_i,r). \qquad (3.15)
\end{aligned}$$

Since (3.13) holds for $T \geq T_0$, the first term of the right-hand side of (3.15) is bounded by

$$\frac{1}{n-m+1}\sum_{t=0}^{T-1}d_k\left(x_{tk+1}^{(t+1)k},\pi_k(x_{tk+1}^{(t+1)k})\right) \leq \left(1+\frac{m}{Tk-m}\right)\left(\delta(x,\pi_k)+\gamma\right). \qquad (3.16)$$

On the other hand, the second term of the right-hand side of (3.15) is bounded as follows. By Lemma 3.2, we have

$$\limsup_{n\to\infty}\frac{1}{n-m+1}\sum_{i=1}^{n}d_1(x_i,r) = \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}d_1(x_i,r) < \infty.$$

Hence, for sufficiently large n and m,

$$\frac{k}{m(n-m+1)}\sum_{i=1}^{n}d_1(x_i,r) \leq \gamma. \qquad (3.17)$$

Thus, if n is so large that $T \geq T_0$, then, combining (3.15), (3.16), and (3.17), we have

$$\begin{aligned}
&\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}d_m\left(x_i^{i+m-1},\pi_m^R(x_i^{i+m-1})\right)\\
&\leq \quad \left(1+\frac{m}{Tk-m}\right)\left(\delta(x,\pi_k)+\gamma\right)+\gamma.
\end{aligned}$$

Since $m/(Tk-m) \to 0$ as $T \to \infty$, we have

$$\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}d_m\left(x_i^{i+m-1},\pi_m^R(x_i^{i+m-1})\right) \leq \delta(x,\pi_k)+3\gamma.$$

Since $\delta(x,\pi_k) \leq \delta(x,\pi_{k'})$ and $\gamma > 0$ can be chosen arbitrarily small, we have (3.12). $\square$

Proof of Theorem 3.2: Fix $\gamma > 0$, and choose $\epsilon > 0$ such that

$$\hat{D}(R-\epsilon|x) \le \lim_{\zeta\downarrow 0} \hat{D}(R-\zeta|x) + \gamma.$$

Then, choose sufficiently large k and n_0 such that $k > 1/\epsilon$. Then, for any $n \ge n_0$,

$$\begin{aligned}
\hat{D}_k(R-\epsilon|x_1^n) &= \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k)\le R-\epsilon}} \frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) \\
&\le \hat{D}(R-\epsilon|x) + \gamma \\
&\le \lim_{\zeta\downarrow 0} \hat{D}(R-\zeta|x) + 2\gamma.
\end{aligned}$$

For each $n \ge n_0$, we can choose a quantizer π_k such that $\mathrm{rate}(\pi_k) \le R-\epsilon$ and

$$\begin{aligned}
&\frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) \\
&\quad\le \lim_{\zeta\downarrow 0} \hat{D}(R-\zeta|x) + 3\gamma, \qquad (3.18)
\end{aligned}$$

where π_k may depend on n. By Lemma 3.4, there exists $m_0 = m_0(\epsilon, k, \mathcal{Y}, L)$ such that for any $m \ge m_0$, we have

$$\begin{aligned}
&d_m\left(x_{tm+1}^{(t+1)m}, \pi_m^R(x_{tm+1}^{(t+1)m})\right) \\
&\le \sum_{i=tm+1}^{(t+1)m-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) + \Delta_k(x_{tm+1}^{(t+1)m}, r),
\end{aligned}$$

for any n and t ($(t+1)m \le n$). It should be noted that m_0 depends on neither π_k nor n. Hence, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{m}d_m\left(x_{tm+1}^{(t+1)m}, \pi_m^R(x_{tm+1}^{(t+1)m})\right) \\
&\le \frac{1}{Tm}\sum_{t=0}^{T-1}\sum_{i=tm+1}^{(t+1)m-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) + \frac{1}{Tm}\sum_{i=0}^{T-1}\Delta_k(x_{tm+1}^{(t+1)m}, r) \\
&= \frac{1}{Tm}\sum_{i=1}^{Tm-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) + \frac{1}{Tm}\sum_{i=0}^{T-1}\Delta_k(x_{tm+1}^{(t+1)m}, r) \\
&\le \frac{1}{Tm-k+1}\sum_{i=1}^{Tm-k+1}\frac{1}{k}d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) \\
&\quad + \frac{1}{Tm}\sum_{i=0}^{T-1}\Delta_k(x_{tm+1}^{(t+1)m}, r). \qquad (3.19)
\end{aligned}$$

Now, we bound the second term of right-hand side of (3.19). By the definition of Δ_k, we have

$$\frac{1}{Tm}\sum_{i=0}^{T-1}\Delta_k(x_{tm+1}^{(t+1)m}, r) = \frac{1}{Tm}\sum_{i\in B} d_1(x_i, r),$$

where

$$B \triangleq \left\{ i: \begin{array}{c} tm+1 \le i \le tm+k-1 \text{ or} \\ (t+1)m-k+2 \le i \le (t+1)m, \\ t = 0, 1, \cdots, T-1 \end{array} \right\}.$$

Since $|B| \le 2(k-1)T$, we can choose m so that $|B|/Tm \le \epsilon$. Thus, by Lemma 3.3, for sufficiently large T and m, we have

$$\frac{1}{Tm}\sum_{i=0}^{T-1}\Delta_k(x_{tm+1}^{(t+1)m}, r) \le \gamma. \tag{3.20}$$

Combining (3.18), (3.19), and (3.20), for sufficiently large T, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{m}d_m\left(x_{tm+1}^{(t+1)m}, \pi_m^R(x_{tm+1}^{(t+1)m})\right) \le \lim_{\zeta\downarrow 0}\hat{D}(R-\zeta|x) + 4\gamma.$$

Since $\gamma > 0$ can be chosen arbitrarily small, we have

$$\limsup_{m\to\infty}\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{m}d_m\left(x_{tm+1}^{(t+1)m}, \pi_m^R(x_{tm+1}^{(t+1)m})\right) \le \lim_{\zeta\downarrow 0}\hat{D}(R-\zeta|x).$$

This completes the proof of the lemma. □

Proof of Theorem 3.3: Since

$$\tilde{D}(R|x) \le \inf_k \inf_{\substack{\pi_k \\ \text{rate}(\pi_k)\le R}} \delta(x, \pi_k) \le \liminf_{k\to\infty}\delta(x, \pi_k^R)$$

holds by the definitions, we shall show

$$\limsup_{k\to\infty}\delta(x, \pi_k^R) \le \lim_{\zeta\downarrow 0}\tilde{D}(R-\zeta|x). \tag{3.21}$$

Fix $\gamma > 0$, and then, choose $\epsilon > 0$, sufficiently large k ($k > 1/\epsilon$), and T_0 such that for any $T \ge T_0$,

$$\inf_{\substack{\pi_k \\ \text{rate}(\pi_k)\le R-\epsilon}} \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) \le \lim_{\zeta\downarrow 0}\tilde{D}(R-\zeta|x) + \gamma.$$

For each $T \geq T_0$, we can choose a quantizer π_k such that rate$(\pi_k) \leq R - \epsilon$ and

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) \leq \lim_{\zeta\downarrow 0}\tilde{D}(R-\zeta|x) + 2\gamma, \tag{3.22}$$

where π_k may depend on T. By Lemma 3.4, there exists $m_0 = m_0(\epsilon, k, \mathcal{Y}, L)$ such that for any $m \geq m_0$ and any $w \in \mathcal{N}$, we have

$$\begin{aligned}
& d_m\left(x_{wm+1}^{(w+1)m}, \pi_m^R(x_{wm+1}^{(w+1)m})\right) \\
& \leq \sum_{\substack{t \\ wm\leq tk\leq (w+1)m-k}} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \Delta_k(x_{wm+1}^{(w+1)m}, r).
\end{aligned}$$

Hence, letting $T = \lfloor Wm/k \rfloor$, we have

$$\begin{aligned}
& \frac{1}{W}\sum_{w=0}^{W-1}\frac{1}{m}d_m\left(x_{wm+1}^{(w+1)m}, \pi_m^R(x_{wm+1}^{(w+1)m})\right) \\
\leq\ & \frac{1}{Wm}\sum_{w=0}^{W-1}\sum_{\substack{t \\ wm\leq tk\leq (w+1)m-k}} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \frac{1}{Wm}\sum_{w=0}^{W-1}\Delta_k(x_{wm+1}^{(w+1)m}, r) \\
=\ & \frac{1}{Wm}\sum_{t=0}^{\lfloor Wm/k\rfloor -1} d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \frac{1}{Wm}\sum_{w=0}^{W-1}\Delta_k(x_{wm+1}^{(w+1)m}, r) \\
\leq\ & \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}d_k\left(x_{tk+1}^{(t+1)k}, \pi_k(x_{tk+1}^{(t+1)k})\right) + \frac{1}{Wm}\sum_{w=0}^{W-1}\Delta_k(x_{wm+1}^{(w+1)m}, r).
\end{aligned} \tag{3.23}$$

If W is sufficiently large, the second term of the right hand side of (3.23) is negligible (see (3.20)), and thus, combining (3.22) and (3.23), we have

$$\frac{1}{W}\sum_{w=0}^{W-1}\frac{1}{m}d_m\left(x_{wm+1}^{(w+1)m}, \pi_m^R(x_{wm+1}^{(w+1)m})\right) \leq \lim_{\zeta\downarrow 0}\tilde{D}(R-\zeta|x) + 3\gamma.$$

Since $\gamma > 0$ can be chosen arbitrarily small, we have (3.21). □

3.5 Fixed-Rate Lossy Coding of Ergodic Sources

In this section, we assume that $\mathcal{X}$ is a standard alphabet and that the distortion measure $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, \infty)$ is a measurable function. For a stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$, let $D(R|\boldsymbol{X})$ be the *distortion-rate function* of $\boldsymbol{X}$ defined as

$$D(R|\boldsymbol{X}) \triangleq \limsup_{n\to\infty} \inf_{Y_1^n : \frac{1}{n}I(X_1^n; Y_1^n)\leq R} \frac{1}{n}Ed_n(X_1^n; Y_1^n),$$

where

$$I(X_1^n; Y_1^n) \triangleq E\left[\frac{P_{Y_1^n|X_1^n}(Y_1^n|X_1^n)}{P_{Y_1^n}(Y_1^n)}\right].$$

It is known that, if $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ is a stationary ergodic source then the optimal distortion level attainable by fixed-rate lossy encoders is equal to $D(R|\boldsymbol{X})$.

Theorem 3.4 (Proposition 2 of [12]) Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source for which there exists $r \in \mathcal{Y}$ such that $E[d_1(X_1, r)] < \infty$. Then, for any $R > 0$, the following two statements hold.

1. (converse part)

   For any sequence $\{\pi_n\}_{n=1}^{\infty}$ of quantizers satisfying $\mathrm{rate}(\pi_n) \leq R$ $(n \in \mathcal{N})$, we have
   $$\liminf_{n\to\infty} \frac{1}{n} d_n\left(x_1^n, \pi_n(x_1^n)\right) \geq D(R|\boldsymbol{X}),$$
   for almost all $x \in \mathcal{X}^{\infty}$.

2. (direct part)

   There exits a sequence $\{\pi_n\}_{n=1}^{\infty}$ of quantizers satisfying $\mathrm{rate}(\pi_n) \leq R$ $(n \in \mathcal{N})$ and
   $$\limsup_{n\to\infty} \frac{1}{n} d_n\left(x_1^n, \pi_n(x_1^n)\right) \leq D(R|\boldsymbol{X}),$$
   for almost all $x \in \mathcal{X}^{\infty}$.

In the following, we consider the blockwise lossy coding of the sequence emitted from the stationary ergodic source. It is important to note that ergodicity of $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ with respect to the shift transformation does *not* imply ergodicity with respect to k-shift transformation. Therefore, we cannot directly apply the ergodic theorem for successive blocks $X_{tk+1}^{(t+1)k}$ $(t = 0, 1, \cdots)$. However, using Theorem 3.2, we can show that the sample average of the block distortion achieved by the quantizer π_k^R based on the complexity function L tends to $D(R|\boldsymbol{X})$.

Theorem 3.5 Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source for which there exists $r \in \mathcal{Y}$ such that $E[d_1(X_1, r)] < \infty$. Then, for any $R > 0$, we have

$$\lim_{k\to\infty} \delta(x, \pi_k^R) = D(R|\boldsymbol{X}), \tag{3.24}$$

for almost all $x \in \mathcal{X}^{\infty}$.

The difference between Theorem 3.4 and Theorem 3.5 should be pointed out here. In the coding scheme appeared in Theorem 3.4, the quantizer quantizes the given whole sequence at one time. On the other hand, in the coding scheme appeared in (3.24), the given sequence is partitioned into blocks of size k, and then, each block is quantized by π_k^R. Theorem 3.5 ensures that the increment of the distortion is negligibly small when we fix the block size k of the quantizer and quantize each block separately, if the block size k is sufficiently large.

Remark 3.3 Yang and Kieffer [31] proved that (3.24) holds when the distortion measure d_n is additive and the complexity function L is the length function L^{LZ} of the LZ78 code.

Before proving Theorem 3.5, we show that if $r \in \mathcal{Y}$ satisfies $E[d_1(X_1, r)] < \infty$, then r satisfies (3.2) for almost all $x \in \mathcal{X}^\infty$.

Lemma 3.5 Let $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ be a stationary ergodic source for which there exists $r \in \mathcal{Y}$ such that

$$E[d_1(X_1, r)] < \infty. \tag{3.25}$$

Then, for almost all $x \in \mathcal{X}^\infty$, we have

$$\lim_{d\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{\substack{1\le i\le n \\ d_1(x_i,r)\ge d}} d_1(x_i, r) = 0.$$

Proof: To prove the lemma, we show that there exists a set $A \subset \mathcal{X}^\infty$ satisfying the following two properties: (i) $P_{X_1^\infty}(A) = 1$ and (ii) for any $x \in A$ and any $\gamma > 0$, there exists $D_0 = D_0(x, \gamma)$ and $n_0 = n_0(x, \gamma)$ such that

$$\frac{1}{n} \sum_{\substack{1\le i\le n \\ d_1(x_i,r)\ge d}} d_1(x_i, r) < \gamma \quad \forall d \ge D_0 \text{ and } \forall n \ge n_0.$$

Fix $\gamma > 0$. Since r satisfies (3.25), we can choose $D_1 = D_1(\boldsymbol{X}, \gamma)$ so that

$$\int_{\{a: d_1(a,r)\ge d\}} d_1(a, r) P_{X_1}(a) < \gamma/2 \quad \forall d \ge D_1.$$

Define the function f_d on $\mathcal{X}$ by

$$f_d(a) \triangleq \begin{cases} d_1(a, r) & : \text{ if } d_1(a, r) \ge d, \\ 0 & : \text{ otherwise.} \end{cases}$$

By the ergodic theorem [5], we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_d(x_i) = E[f_d(X_1)] \quad \text{a.s.}$$

that is,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\substack{1\le i\le n\\ d_1(x_i,r)\ge d}} d_1(x_i,r) = \int_{\{a:d_1(a,r)\ge d\}} d_1(a,r)P_{X_1}(a) \quad \text{a.s.}$$

Hence, for almost all $x \in \mathcal{X}^\infty$, we can choose $n_1 = n_1(x,\gamma)$ so that

$$\frac{1}{n}\sum_{\substack{1\le i\le n\\ d_1(x_i,r)\ge d}} d_1(x_i,r) < \gamma \quad \forall d \ge D_1 \text{ and } \forall n \ge n_1.$$

This implies that for any $\gamma > 0$, there exists $A_\gamma \subset \mathcal{X}^\infty$ satisfying $P_{X_1^\infty}(A_\gamma) = 1$ and

$$\lim_{d\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{\substack{1\le i\le n\\ d_1(x_i,r)\ge d}} d_1(x_i,r) < \gamma \quad \forall x \in A_\gamma.$$

Let $A = \cap_{m=1}^\infty A_{1/m}$. Then, $A_{1/1} \supset A_{1/2} \supset \cdots$ and thus,

$$\begin{aligned} P_{X_1^\infty}(A) &= P_{X_1^\infty}(\cap_{m=1}^\infty A_{1/m}) \\ &= \lim_{m\to\infty} P_{X_1^\infty}(A_{1/m}) \\ &= 1. \end{aligned}$$

Note that for any $\gamma > 0$, there exists $m \in \mathcal{N}$ such that $1/m < \gamma$. Hence, for all $x \in A$ and any $\gamma > 0$, there exists $D_0 = D_0(x,\gamma) = D_1(\boldsymbol{X}, 1/m)$ and $n_0 = n_0(x,\gamma) = n_1(x, 1/m)$ such that

$$\frac{1}{n}\sum_{\substack{1\le i\le n\\ d_1(x_i,r)\ge d}} d_1(x_i,r) < \gamma \quad \forall d \ge D_0 \text{ and } \forall n \ge n_0.$$

Hence, we have the lemma. □

Proof of Theorem 3.5: The converse part of Theorem 3.4 implies that

$$\liminf_{k\to\infty} \delta(x, \pi_k^R) \ge D(R|\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$. Hence, to show the theorem, it is sufficient to show that for almost all $x \in \mathcal{X}^\infty$,

$$\limsup_{k\to\infty} \delta(x, \pi_k^R) \le D(R|\boldsymbol{X}). \tag{3.26}$$

Since if $D(R|\boldsymbol{X}) = \infty$ then (3.26) holds, we assume that $D(R|\boldsymbol{X}) < \infty$. Then, for almost all $x \in \mathcal{X}^\infty$, we have

$$
\begin{aligned}
&\hat{D}(R|x) \\
&\leq \liminf_{k\to\infty} \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k)\leq R}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{k} d_k\left(x_i^{i+k-1}, \pi_k(x_i^{i+k-1})\right) && (3.27)\\
&= \liminf_{k\to\infty} \inf_{\substack{\pi_k \\ \mathrm{rate}(\pi_k)\leq R}} E\left[\frac{1}{k} d_k\left(X_1^k, \pi_k(X_1^k)\right)\right] && (3.28)\\
&= D(R|\boldsymbol{X}) && (3.29)
\end{aligned}
$$

where (3.27) follows from the definition, (3.28) follows from the ergodic theorem [5], and (3.29) follows from the direct part of Theorem 3.4.

On the other hand, Lemma 3.5 implies that we can apply Theorem 3.2 for almost all $x \in \mathcal{X}^\infty$. Thus, by Theorem 3.2 and (3.29), for almost all $x \in \mathcal{X}^\infty$, we have

$$
\begin{aligned}
\limsup_{k\to\infty} \delta(x, \pi_k^R) &\leq \lim_{\zeta\downarrow 0} \hat{D}(R-\zeta|x) \\
&\leq \lim_{\zeta\downarrow 0} D(R-\zeta|\boldsymbol{X}).
\end{aligned}
$$

Since $D(\cdot|\boldsymbol{X})$ is continuous [6], we have (3.26). □

Chapter 4

Fixed-Distortion Lossy Coding Problem

4.1 Introduction

In Chapter 3, we considered the fixed-rate lossy coding problem of individual sequences. In this chapter, we consider the variable-rate lossy coding of individual sequences, and investigate the optimal compression rate at the fixed-distortion level.

In a similar manner as the fixed-rate coding problem, we formulate the fixed-distortion lossy coding problem of individual sequences as the problem of optimizing the coding rate over all the blockwise lossy encoders. Then, we characterize the OPTA function of the given individual sequence by using the empirical distribution of the sequence. Our result clarifies that the optimal coding rate achievable by fixed-distortion lossy coding schemes is equal to the optimal average coding rate with respect to the overlapping empirical distribution of the given sequence. We also investigate relations between the OPTA function and the rate-distortion function derived by Yang and Kieffer [31]. In addition, we propose the fixed-distortion universal lossy coding scheme based on the complexity of the sequence.

4.2 Fixed-Distortion Lossy Coding of Individual Sequences

Before stating our result, we introduce some assumptions. In this chapter, we assume that we can choose a distortion level D satisfying

$$D > \sup_{a \in \mathcal{X}} \min_{b \in \mathcal{Y}} d_1(a, b). \tag{4.1}$$

In addition, unless otherwise specified, we assume that the distortion measure satisfies that

$$\frac{1}{m+n} d_{m+n}(x_1^{m+n}, y_1^{m+n}) \leq \max \left\{ \frac{1}{m} d_m(x_1^m, y_1^m), \frac{1}{n} d_n(x_{m+1}^{m+n}, y_{m+1}^{m+n}) \right\}. \tag{4.2}$$

As shown in the following lemma, if the distortion measure is subadditive then (4.2) holds. On the other hand, there exists a distortion measure satisfying (4.2) that are not subadditive, for example,

$$d_n(x_1^n, y_1^n) = n \max_{1 \leq i \leq n} d_1(x_i, y_i),$$

where $d_1 : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$.

Lemma 4.1 If the distortion measure d_n is subadditive, then (4.2) holds.

Proof: Without loss of generality, we can assume that

$$\frac{1}{m} d_m(x_1^m, y_1^m) \geq \frac{1}{n} d_n(x_{m+1}^{m+n}, y_{m+1}^{m+n}).$$

This implies that

$$d_m(x_1^m, y_1^m) + d_n(x_{m+1}^{m+n}, y_{m+1}^{m+n}) \leq \frac{m+n}{m} d_m(x_1^m, y_1^n). \tag{4.3}$$

Combining (3.1) and (4.3), we have

$$\frac{1}{m+n} d_{m+n}(x_1^{m+n}, y_1^{m+n}) \leq \frac{1}{m} d_m(x_1^m, y_1^m).$$

□

Now, we formulate the fixed-distortion lossy coding problem of individual sequences. A variable-rate lossy encoder of block size k is defined by a quantizer $\pi_k : \mathcal{X}^k \to \mathcal{Y}^k$ and a length function σ_k on $\mathcal{Y}^k$. By a given encoder

$$\begin{array}{cccc}
\underbrace{x_1^k} & \underbrace{x_{k+1}^{2k}} & \cdots & \underbrace{x_{(t-1)k+1}^{tk}} \\
\downarrow & \downarrow & & \downarrow \\
\boxed{\pi_k} & \boxed{\pi_k} & \cdots & \boxed{\pi_k} \\
\downarrow & \downarrow & & \downarrow \\
\underbrace{y_1^k} & \underbrace{y_{k+1}^{2k}} & \cdots & \underbrace{y_{(t-1)k+1}^{tk}} \\
\downarrow & \downarrow & & \downarrow \\
\boxed{\sigma_k} & \boxed{\sigma_k} & \cdots & \boxed{\sigma_k} \\
\downarrow & \downarrow & & \downarrow
\end{array}$$

Figure 4.1: The blockwise lossy encoder (π_k, σ_k)

(π_k, σ_k), a sequence $x \in \mathcal{X}^\infty$ is encoded in the following manner. At first, x is partitioned into blocks of size k, and then, each block $x_{tk+1}^{(t+1)k}$ $(t = 0, 1, 2, \cdots)$ is quantized into $y_{tk+1}^{(t+1)k} = \pi_k(x_{(t-1)k+1}^{tk})$ by π_k. Then, each $y_{tk+1}^{(t+1)k}$ is encoded by σ_k (see Fig. 4.1). The asymptotic compression rate $\rho(x, \pi_k, \sigma_k)$ attained by the variable-rate lossy encoder (π_k, σ_k) is defined as

$$\rho(x, \pi_k, \sigma_k) \triangleq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{k} \sigma_k \left(\pi_k(x_{tk+1}^{(t+1)k}) \right).$$

To simplify the notation, we write $\mathrm{dist}(\pi_k) \le D$ meaning that the quantizer π_k satisfies that $d_k(a_1^k, \pi_k(a_1^k)) \le kD$ for all $a_1^k \in \mathcal{X}^k$. Then, the optimal compression rate $\rho(D|x)$ attainable by blockwise lossy coding at the distortion level D is defined as

$$\rho(D|x) \triangleq \inf_k \inf_{\substack{(\pi_k, \sigma_k) \\ \mathrm{dist}(\pi_k) \le D}} \rho(x, \pi_k, \sigma_k),$$

where $\inf_{\substack{(\pi_k, \sigma_k) \\ \mathrm{dist}(\pi_k) \le D}}$ is taken over all (π_k, σ_k) satisfying $\mathrm{dist}(\pi_k) \le D$ (Since (4.1) and (4.2) hold, there exists at least one quantizer π_k such that $\mathrm{dist}(\pi_k) \le D$).

In order to clarify the optimal compression rate $\rho(D|x)$, we introduce the rate-distortion function for each individual sequence. Let $p_k(\cdot|x_1^n)$ be the overlapping empirical distribution of x_1^n defined in (2.1). Then, let $\hat{R}_k(D|x_1^n)$ be the optimal average compression rate with respect to $p_k(\cdot|x_1^n)$:

$$\hat{R}_k(D|x_1^n) \triangleq \inf_{\substack{(\pi_k, \sigma_k) \\ \mathrm{dist}(\pi_k) \le D}} E_p \left[\frac{1}{k} \sigma_k \left(\pi_k(X_1^k) \right) \right]$$

$$= \inf_{\substack{(\pi_k,\sigma_k) \\ \mathrm{dist}(\pi_k)\le D}} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \frac{1}{k} \sigma_k \left(\pi_k(x_i^{i+k-1}) \right).$$

Lastly, the *rate-distortion function* $\hat{R}(D|x)$ of the sequence $x \in \mathcal{X}^\infty$ is defined as

$$\hat{R}(D|x) \triangleq \liminf_{k\to\infty} \limsup_{n\to\infty} \hat{R}_k(D|x_1^n).$$

The following theorem gives a lower bound on the compression rate attainable by variable-rate block encoders.

Theorem 4.1 (Converse coding theorem) Assume that D satisfies (4.1). For any variable-rate lossy encoder (π_k, σ_k) of the block size k satisfying $\mathrm{dist}(\pi_k) \le D$, we have

$$\rho(x, \pi_k, \sigma_k) \ge \hat{R}(D|x) \quad \forall x \in \mathcal{X}^\infty. \tag{4.4}$$

Next, we show the coding theorem by constructing a universal fixed-rate lossy block encoder. To describe the universal lossy block encoder, we first define the quantizer based on the complexity function L. For each $k = 1, 2, \cdots$, let π_k^D be the quantizer defined as

$$\pi_k^D(a_1^k) \triangleq \arg \min_{\substack{b_1^k \in \mathcal{Y}^k \\ d_k(a_1^k, b_1^k) \le kD}} L(b_1^k) \quad \forall a_1^k \in \mathcal{X}^k. \tag{4.5}$$

Since L satisfies the Kraft inequality (3.4), the pair (π_k^D, L) defines the variable-rate lossy encoder of block size k.

Now, we state the coding theorem.

Theorem 4.2 (Coding theorem) Assume that D satisfies (4.1). The sequence $\{(\pi_k^D, L)\}_{k=1}^\infty$ of lossy block encoders satisfies that

$$\limsup_{k\to\infty} \rho(x, \pi_k^D, L) \le \hat{R}(D|x) \quad \forall x \in \mathcal{X}^\infty. \tag{4.6}$$

Theorem 4.1 and Theorem 4.2 will be proved in Section 4.4. According to Theorem 4.1 and Theorem 4.2, we have the following corollary.

Corollary 4.1 For any $x \in \mathcal{X}^\infty$ and D satisfying (4.1), we have

$$\lim_{k\to\infty} \rho(x, \pi_k^D, L) = \rho(D|x) = \hat{R}(D|x).$$

4.3 Relations among Rate-Distortion Functions of Individual Sequence

As in the case of the distortion-rate function, the rate-distortion function may be defined by using the non-overlapping empirical distribution instead of the overlapping empirical distribution. Let $\tilde{R}_k(D|x_1^{Tk})$ be the optimal average compression rate with respect to $q_k(\cdot|x_1^{Tk})$,

$$\begin{aligned}\tilde{R}_k(D|x_1^{Tk}) &\triangleq \inf_{\substack{(\pi_k,\sigma_k)\\ \mathrm{dist}(\pi_k)\le D}} E_q\left[\frac{1}{k}\sigma_k\left(\pi_k(X_1^k)\right)\right] \\ &= \inf_{\substack{(\pi_k,\sigma_k)\\ \mathrm{dist}(\pi_k)\le D}} \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right),\end{aligned}$$

and $\tilde{R}(D|x)$ be

$$\tilde{R}(D|x) \triangleq \liminf_{k\to\infty}\limsup_{T\to\infty}\tilde{R}_k(D|x_1^{Tk}).$$

Now, we show that $\tilde{R}(D|x)$ is identical with the rate-distortion function $R^{YK}(D|x)$ introduced by Yang and Kieffer [31]. For x_1^{Tk} and π_k, let $q_k(\cdot|x_1^{Tk},\pi_k)$ be the non-overlapping empirical distribution defined as

$$q_k(b_1^k|x_1^{Tk},\pi_k) \triangleq \frac{|\{t : 0\le t\le T-1, \pi_k(x_{tk+1}^{(t+1)k}) = b_1^k\}|}{T},$$

and $\tilde{H}_k^{Tk}(x_1^{Tk},\pi_k)$ be the entropy of $q_k(\cdot|x_1^{Tk},\pi_k)$ defined by

$$\tilde{H}_k^n(x_1^n,\pi_k) \triangleq \frac{1}{k}\sum_{b_1^k\in\mathcal{Y}^k} -q_k(b_1^k|x_1^n,\pi_k)\log q_k(b_1^k|x_1^n,\pi_k).$$

For a fixed quantizer π_k, let σ_k be the length function of Huffman code (see, for example, [2]) for $q_k(\cdot|x_1^{Tk},\pi_k)$. Then, we have

$$\begin{aligned}\tilde{H}_k^{Tk}(x^{Tk},\pi_k) &\le \inf_{\sigma_k}\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right) \\ &\le \tilde{H}_k^{Tk}(x^{Tk},\pi_k) + \frac{1}{Tk}.\end{aligned}$$

Hence, we have

$$\tilde{R}(D|x) = \liminf_{k\to\infty}\limsup_{T\to\infty}\inf_{\substack{\pi_k\\ \mathrm{dist}(\pi_k)\le D}}\tilde{H}_k^{Tk}(x_1^{Tk},\pi_k), \tag{4.7}$$

and the right hand side is identical with the definition of $R^{YK}(D|x)$ [31].

On the other hand, Yang and Kieffer [31] proposed a lossy encoder (π_k^{LZ}, L^{LZ}) based on the length function L^{LZ} of LZ78 code, and showed that

$$\lim_{k\to\infty} \rho(x, \pi_k^{LZ}, L^{LZ}) = R^{YK}(D|x), \tag{4.8}$$

under the condition that the distortion measure is additive. Since L^{LZ} is a complexity function, combining Corollary 4.1, (4.7), and (4.8), we have the following corollary. This corollary shows that $R^{YK}(D|x)$ is equal to $\hat{R}(D|x)$ and $\tilde{R}(D|x)$ if the distortion measure is additive.

Corollary 4.2 Assume that the distortion measure d_n is additive. Then, for any $x \in \mathcal{X}^\infty$, we have

$$\hat{R}(D|x) = \tilde{R}(D|x) = R^{YK}(D|x).$$

In [17], the rate-distortion function $R^*_{vm}(D|x)$ of x was defined as the infimum of R such that there exists $y \in \mathcal{Y}^\infty$ satisfying the distortion constraint

$$\limsup_{k\to\infty} \sup_i \frac{1}{k} d_k(x_i^{i+k-1}, y_i^{i+k-1}) \le D, \tag{4.9}$$

and the rate constraint

$$\hat{H}(y) \le R,$$

where $\hat{H}(y)$ is the overlapping empirical entropy of y. As shown in Theorem 2.1, $\hat{H}(y)$ is the finite-state compressibility $\rho(y)$, that is, the optimal compression rate of y attainable by finite-state lossless encoders. We can interpret the meaning of $R^*_{vm}(D|x)$ as the infimum of the optimal lossless compression rate $\hat{H}(y)$ of the reproduction sequence y satisfying the distortion constraint (4.9). The following theorem shows that $R^*_{vm}(D|x)$ is equal to the rate-distortion function $\hat{R}(D|x)$ if the distortion measure is non-decreasing.

Theorem 4.3 Assume that the distortion measure d_n satisfies (4.2) and is non-decreasing, that is, for any $n \in \mathcal{N}$, $x_1^n \in \mathcal{X}^n$, and $y_1^n \in \mathcal{Y}^n$,

$$d_k(x_i^{i+k-1}, y_i^{i+k-1}) \le d_n(x_1^n, y_1^n) \quad \forall k, i \ (1 \le i \le n-k+1).$$

Then, we have

$$\limsup_{k\to\infty} \rho(x, \pi_k^D, L) = \hat{R}(D|x) = R^*_{vm}(D|x) \quad x \in \mathcal{X}^\infty,$$

provided that $R^*_{vm}(\cdot|x)$ is continuous at the point D.

Theorem 4.3 will be proved in Section 4.4.

4.4 Proofs of Theorems

In this section, we prove Theorem 4.1, Theorem 4.2, and Theorem 4.3.

At first, we prove a key lemma. Using this lemma, Theorem 4.1 (resp. Theorem 4.2) can be proved by a similar manner as the proof of Theorem 3.1 (resp. Theorem 3.2).

Lemma 4.2 Let L be a complexity function, and assume that D satisfies (4.1). For any $k \in \mathcal{N}$ and any $\gamma > 0$, there exists $m_0 = m_0(\gamma, k, \mathcal{Y}, L)$ such that for any (π_k, σ_k) satisfying $\mathrm{dist}(\pi_k) \leq D$ and any $m \geq m_0$, the quantizer π_m^D defined in (4.5) satisfies that

$$L\left(\pi_m^D(a_1^m)\right) \leq \min_{0\leq j\leq k-1} \sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} \sigma_k\left(\pi_k(a_{i+1}^{i+k})\right) + m\gamma \quad \forall a_1^m \in \mathcal{X}^m, \tag{4.10}$$

and thus,

$$L\left(\pi_m^D(a_1^m)\right) \leq \sum_{i=0}^{m-k} \frac{1}{k}\sigma_k\left(\pi_k(a_{i+1}^{i+k})\right) + m\gamma \quad \forall a_1^m \in \mathcal{X}^m. \tag{4.11}$$

Proof: Let $\pi_1^\star$ be the quantizer defined as

$$\pi_1^\star(a) \stackrel{\triangle}{=} \arg\min_{b\in\mathcal{Y}} d_1(a,b).$$

Fix j $(0 \leq j \leq k-1)$ and let $\bar{j} \stackrel{\triangle}{=} \max\{l : l \equiv j \bmod k, 1 \leq l \leq m\}$. Further, let $b(j)_1^m$ be the block of length m defined as

$$b(j)_i \stackrel{\triangle}{=} \pi_1^\star(a_i) \quad 1 \leq i \leq j \text{ or } \bar{j}+1 \leq i \leq m,$$

and

$$b(j)_{i+1}^{i+k} = \pi_k(a_{i+1}^{i+k}) \quad i \equiv j \bmod k \text{ and } 0 \leq i \leq m-k$$

(see Fig. 4.2). According to (4.1), $\pi_1^\star$ satisfies

$$d_1(a, \pi_1^\star(a)) < D \quad \forall a \in \mathcal{X}.$$

Moreover, since $\mathrm{dist}(\pi_k) \leq D$ and (4.2) holds, we have

$$d_m(a_1^m, b(j')_1^m) \leq mD \quad \forall j' = 0, \cdots, k-1. \tag{4.12}$$

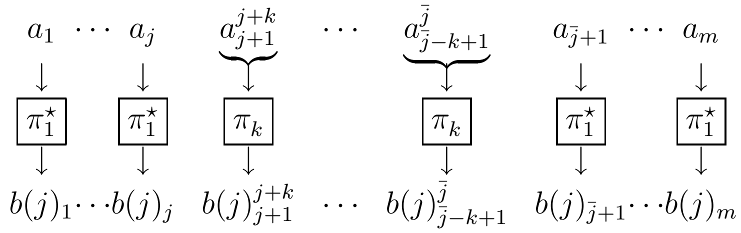

Figure 4.2: Definition of $b(j)_1^m$

Thus, for sufficiently large m, we have

$$
\begin{aligned}
L\left(\pi_m^D(a_1^m)\right) &\le \min_{0\le j'\le k-1} L(b(j')_1^m) && (4.13)\\
&\le \min_{0\le j'\le k-1}\min_{0\le j\le k-1}\sum_{\substack{0\le i\le m-k\\ i\equiv j \bmod k}} \sigma_k\left(b(j')_{i+1}^{i+k}\right)+m\gamma && (4.14)\\
&\le \min_{0\le j'\le k-1}\sum_{\substack{0\le i\le m-k\\ i\equiv j' \bmod k}} \sigma_k\left(b(j')_{i+1}^{i+k}\right)+m\gamma\\
&= \min_{0\le j\le k-1}\sum_{\substack{0\le i\le m-k\\ i\equiv j \bmod k}} \sigma_k\left(\pi_k(a_{i+1}^{i+k})\right)+m\gamma,
\end{aligned}
$$

where (4.13) follows from (4.12) and the definition of π_m^D, and (4.14) from Lemma 3.1. Hence, we have the lemma. $\square$

Proof of Theorem 4.1: Notice that

$$
\begin{aligned}
&\hat{R}(D|x)\\
&= \liminf_{m\to\infty}\limsup_{n\to\infty}\inf_{\substack{(\pi_m,\sigma_m)\\ \mathrm{dist}(\pi_m)\le D}} \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}\sigma_m\left(\pi_m(x_i^{i+m-1})\right)\\
&\le \limsup_{m\to\infty}\limsup_{n\to\infty}\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right).
\end{aligned}
$$

Hence, to show the theorem, it is sufficient to show that

$$
\limsup_{m\to\infty}\limsup_{n\to\infty}\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right)\le\rho(x,\pi_k,\sigma_k) \qquad (4.15)
$$

for any (π_k,σ_k) such that $\mathrm{dist}(\pi_k)\le D$.

Fix $\gamma > 0$ and choose T_0 such that for any $T \geq T_0$,

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right) \leq \rho(x,\pi_k,\sigma_k)+\gamma. \tag{4.16}$$

By Lemma 4.2, there exists $m_0 = m_0(\gamma, k, \mathcal{Y}, L)$ such that for any $m \geq m_0$,

$$L\left(\pi_m^D(x_i^{i+m-1})\right) \leq \min_{0\leq j\leq k-1}\sum_{\substack{l\equiv j \bmod k\\ 0\leq l\leq m-k}}\sigma_k\left(\pi_k(x_{i+l}^{i+l+k-1})\right)+m\gamma,$$

for any $i \in \mathcal{N}$. This implies that

$$\begin{aligned} L\left(\pi_m^D(x_i^{i+m-1})\right) &\leq \sum_{\substack{l\equiv j_i \bmod k\\ 0\leq l\leq m-k}}\sigma_k\left(\pi_k(x_{i+l}^{i+l+k-1})\right)+m\gamma,\\ &= \sum_{\substack{t\\ i\leq(tk+1)\leq i+m-k}}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right)+m\gamma, \end{aligned}$$

where j_i is chosen so that $i + j_i \equiv 1 \bmod k$. Hence, we have

$$\begin{aligned} &\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right)\\ &\leq \frac{1}{m}\sum_{i=1}^{n-m+1}\sum_{\substack{t\\ i\leq(tk+1)\leq i+m-k}}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right)+(n-m+1)\gamma\\ &\leq \sum_{t=0}^{\lfloor n/k\rfloor-1}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right)+(n-m+1)\gamma, \end{aligned} \tag{4.17}$$

where (4.17) follows from the fact that each k-blocks is encoded by (π_k, σ_k) at most m times. Since (4.17) holds for any n and T, letting $T = \lfloor n/k \rfloor$, we have

$$\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right) \leq \frac{1}{n-m+1}\sum_{t=0}^{T-1}\sigma_k\left(\pi_k(x_{tk+1}^{(t+1)k})\right)+\gamma.$$

Since (4.16) holds for $T \geq T_0$, we have

$$\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right) \leq \left(1+\frac{m}{Tk-m}\right)(\rho(x,\pi_k,\sigma_k)+\gamma)+\gamma,$$

for $T \geq T_0$. Since $m/(Tk - m) \to 0$ as $T \to \infty$, we have

$$\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}L\left(\pi_m^D(x_i^{i+m-1})\right) \leq \rho(x,\pi_k,\sigma_k)+3\gamma.$$

Since we can choose $\gamma > 0$ arbitrarily small, we have (4.15). $\square$

Proof of Theorem 4.2: Fix $\gamma > 0$, and choose k and n_0 such that for any $n \geq n_0$,

$$\inf_{\substack{(\pi_k,\sigma_k)\\ \mathrm{dist}(\pi_k)\leq D}} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right) \leq \hat{R}(D|x)+\gamma.$$

For each $n \geq n_0$, we can choose an encoder (π_k,σ_k) such that $\mathrm{dist}(\pi_k) \leq D$ and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right) \leq \hat{R}(D|x)+2\gamma, \tag{4.18}$$

where (π_k,σ_k) may depend on n. By Lemma 4.2, there exists $m_0 = m_0(\gamma,k,\mathcal{Y},L)$ such that for any $m \geq m_0$, we have

$$L\left(\pi_m^D(x_{tm+1}^{(t+1)m})\right) \leq \sum_{i=tm+1}^{(t+1)m-k+1}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right)+m\gamma,$$

for any n and t ($(t+1)m \leq n$). It should be noted that m_0 depends on neither (π_k,σ_k) nor n. Further, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{m}L\left(\pi_m^D(x_{tm+1}^{(t+1)m})\right) &\leq \frac{1}{Tm}\sum_{t=0}^{T-1}\sum_{i=tm+1}^{(t+1)m-k+1}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right)+\gamma\\ &= \frac{1}{Tm}\sum_{i=1}^{Tm-k+1}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right)+\gamma\\ &\leq \frac{1}{Tm-k+1}\sum_{i=1}^{Tm-k+1}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right)+\gamma.\end{aligned}$$

Since (4.18) holds for $n \geq n_0$, and $\gamma(>0)$ can be chosen arbitrarily small, we have

$$\limsup_{m\to\infty}\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{m}L\left(\pi_m^D(x_{tm+1}^{(t+1)m})\right) \leq \hat{R}(D|x).$$

□

Proof of Theorem 4.3: By Corollary 4.1, to show the theorem, it is sufficient to show that

$$\limsup_{k\to\infty}\rho(x,\pi_k^D,L) = R_{vm}^*(D|x) \tag{4.19}$$

provided that $R_{vm}^*(\cdot|x)$ is continuous at the point D.

Let $\bar{y} = \pi_k^D(x) \equiv \pi_k^D(x_1^k)\pi_k^D(x_{k+1}^{2k})\cdots$, and fix $\epsilon > 0$ arbitrarily. For any integer m and i such that $m = l_1k + j_1$ $(1 \leq j_1 \leq k)$ and $i = l_2k + j_2$

$(1 \leq j_2 \leq k)$, we have

$$\begin{aligned} d_m(x_i^{i+m-1}, \bar{y}_i^{i+m-1}) &= d_{l_1k+j_1}(x_{l_2k+j_2}^{(l_1+l_2)k+(j_1+j_2)-1}, \bar{y}_{l_2k+j_2}^{(l_1+l_2)k+(j_1+j_2)-1}) \\ &\leq d_{(l_1+2)k}(x_{l_2k+1}^{(l_1+l_2+2)k}, \bar{y}_{l_2k+1}^{(l_1+l_2+2)k}) \qquad (4.20) \\ &\leq (l_1+2)kD \qquad (4.21) \\ &= m\left(D + \frac{2k-j_1}{l_1k+j_1}D\right), \end{aligned}$$

where (4.20) follows from the fact that d_m is non-decreasing, and (4.21) from $\mathrm{dist}(\pi_k^D) \leq D$. Thus, if $m = l_1k + j_1$ is sufficiently large,

$$d_m(x_i^{i+m-1}, \bar{y}_i^{i+m-1}) \leq m(D+\epsilon) \quad \forall i \in \mathcal{N}.$$

Since ϵ can be chosen arbitrarily small, we have

$$\limsup_{m\to\infty} \sup_i \frac{1}{m} d_m(x_i^{i+m-1}, \bar{y}_i^{i+m-1}) \leq D. \qquad (4.22)$$

Further, by Theorem 2.2, we have

$$\hat{H}(\bar{y}) = \rho^{bl}(\bar{y}) \leq \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{k} L\left(\bar{y}_{(t-1)k+1}^{tk}\right). \qquad (4.23)$$

By the definition of R^*_{vm}, (4.22) and (4.23), we have

$$R^*_{vm}(D|x) \leq \limsup_{k\to\infty} \rho(x, \pi_k^D, L). \qquad (4.24)$$

Next, we show that if $R^*_{vm}(\cdot|x)$ is continuous then

$$\limsup_{k\to\infty} \rho(x, \pi_k^D, L) \leq R^*_{vm}(D|x). \qquad (4.25)$$

Fix $\epsilon > 0$ and $\gamma > 0$ arbitrarily. By the definition of $R^*_{vm}(D-\gamma|x)$, there exists a sequence $y \in \mathcal{Y}^\infty$ and an integer k_0 such that

$$\hat{H}(y) \leq R^*_{vm}(D-\gamma|x) + \epsilon, \qquad (4.26)$$

and for all $k \geq k_0$ and $\forall i \in \mathcal{N}$,

$$\frac{1}{k} d_k(x_i^{i+k-1}, y_i^{i+k-1}) \leq (D-\gamma) + \gamma = D. \qquad (4.27)$$

On the other hand, by Theorem 2.2, for sufficiently large m and T, there exists a length function σ_m^T on $\mathcal{Y}^m$ such that

$$\frac{1}{T}\sum_{i=0}^{T-1} \frac{1}{m} \sigma_m^T(y_{im+1}^{(i+1)m}) \leq \hat{H}(y) + \epsilon/2.$$

By Lemma 3.1, for sufficiently large k ($k \geq m$), we have for all $t = 0, 1, 2, \cdots, T$,

$$L(y_{tk+1}^{(t+1)k}) \leq \sum_{\substack{i: \\ tk \leq im \leq (t+1)k-m}} \sigma_m^T(y_{im+1}^{(i+1)m}) + k\epsilon/2,$$

and thus, for sufficiently large k and T, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}L(y_{tk+1}^{(t+1)k}) \leq \hat{H}(y) + \epsilon. \tag{4.28}$$

Moreover, according to the definition of π_k^D and (4.27), for sufficiently large k ($k \geq k_0$), we have

$$L(\pi_k^D(x_{tk+1}^{(t+1)k})) \leq L(y_{tk+1}^{(t+1)k}). \tag{4.29}$$

Combining (4.26), (4.28) and (4.29), for sufficiently large k and T, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}L\left(\pi_k^D(x_{tk+1}^{(t+1)k})\right) &\leq R_{vm}^*(D-\gamma|x) + 2\epsilon \\ &\leq R_{vm}^*(D|x) + 3\epsilon, \qquad (4.30)\end{aligned}$$

where the last inequality (4.30) follows from the assumption that $R_{vm}^*(\cdot|x)$ is continuous at the point D. Since $\epsilon > 0$ can be chosen arbitrarily small, we have (4.25).

Combining (4.24) and (4.25), we have (4.19) □

4.5 Fixed-Distortion Coding of Ergodic Sources

In this section, we assume that $\mathcal{X}$ is a standard alphabet and that the distortion measure $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, \infty)$ is a measurable function. For a stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$, let $D(R|\boldsymbol{X})$ be the *rate-distortion function* of $\boldsymbol{X}$ defined as

$$R(D|\boldsymbol{X}) \stackrel{\triangle}{=} \inf_n \inf_{Y_1^n:\Pr\{d_n(X_1^n, Y_1^n) \leq nD\}=1} \frac{1}{n}I(X_1^n; Y_1^n).$$

It is known that, if $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ is a stationary ergodic source then the optimal compression rate attainable by the fixed-distortion lossy encoders is equal to $D(R|\boldsymbol{X})$.

Theorem 4.4 (Proposition 4 of [12]) Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source, and assume that D satisfies (4.1). Then, the following two statements hold.

1. (converse part)

   For any sequence $\{(\pi_n, \sigma_n)\}_{n=1}^{\infty}$ of variable-rate encoders satisfying $\mathrm{dist}(\pi_n) \leq D$ $(n \in \mathcal{N})$, we have

$$\liminf_{n\to\infty} \frac{1}{n}\sigma_n\left(\pi_n(x_1^n)\right) \geq R(D|\boldsymbol{X}),$$

   for almost all $x \in \mathcal{X}^\infty$.

2. (direct part)

   There exits a sequence $\{(\pi_n, \sigma_n)\}_{n=1}^{\infty}$ of variable-rate encoders satisfying $\mathrm{rate}(\pi_n) \leq D$ $(n \in \mathcal{N})$ and

$$\limsup_{n\to\infty} \frac{1}{n}\sigma_n\left(\pi_n(x_1^n)\right) \leq R(D|\boldsymbol{X}),$$

   for almost all $x \in \mathcal{X}^\infty$.

In the coding scheme appeared in Theorem 4.4, the given whole sequence is encoded at one time. On the other hand, in the following, we consider the blockwise lossy coding of the sequence emitted from the stationary ergodic source. Next theorem shows that the optimal performance can be attained even if we encode each block separately.

Theorem 4.5 Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source, and assume that D satisfies (4.1). Then, we have

$$\lim_{k\to\infty} \rho(x, \pi_k^D, L) = R(D|\boldsymbol{X}), \tag{4.31}$$

for almost all $x \in \mathcal{X}^\infty$.

Proof: The converse part of Theorem 4.4 implies that for almost all $x \in \mathcal{X}^\infty$,

$$\liminf_{k\to\infty} \rho(x, \pi_k^D, L) \geq R(D|\boldsymbol{X}).$$

Moreover, we have, for almost all $x \in \mathcal{X}^\infty$,

$$\begin{aligned}
&\lim_{k\to\infty} \rho(x, \pi_k^D, L) \\
&\quad= \hat{R}(D|x) && (4.32)\\
&\quad= \liminf_{k\to\infty} \limsup_{n\to\infty} \inf_{\substack{(\pi_k,\sigma_k) \\ \mathrm{dist}(\pi_k)\leq D}} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right) && (4.33)
\end{aligned}$$

$$
\begin{aligned}
&\leq \liminf_{k\to\infty} \inf_{\substack{(\pi_k,\sigma_k)\\ \mathrm{dist}(\pi_k)\leq D}} \limsup_{n\to\infty} \frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}\sigma_k\left(\pi_k(x_i^{i+k-1})\right) \\
&= \liminf_{k\to\infty} \inf_{\substack{(\pi_k,\sigma_k)\\ \mathrm{dist}(\pi_k)\leq D}} E\left[\frac{1}{k}\sigma_k\left(\pi_k(X_1^k)\right)\right] && (4.34)\\
&= R(D|\boldsymbol{X}), && (4.35)
\end{aligned}
$$

where (4.32) follows from Corollary 4.1, (4.33) from the definition of $\hat{R}$, (4.34) from the ergodic theorem [5], and (4.35) from the direct part of Theorem 4.4. Hence, we have the theorem. □

Chapter 5

Fixed-Slope Lossy Coding Problem

5.1 Introduction

In Chapter 3 and Chapter 4, we considered the fixed-rate and fixed-distortion lossy coding problems of individual sequences, and proposed two kinds of universal lossy encoders based on the complexity function. While these universal coding schemes are conceptually simple, it is not easy to implement them because they need a search over a set of reproduction sequences (see (3.6) and (4.5)).

Seeking a universal lossy encoder with low coding complexity, Yang, Zhang, and Berger [33] considered the coding problem called the fixed-slope coding problem. The fixed-slope coding problem considers the minimization of the cost defined as

$$\sigma_n(y_1^n) + \lambda d_n(x_1^n, y_1^n),$$

for some fixed $\lambda > 0$. In this lossy coding problem, neither rate nor distortion needs to be fixed, and thus, to minimize the cost, we can use some sequential search algorithms such as trellis search [33, 32]. Yang *et al.* [33] clarified the optimal cost attainable by fixed-slope lossy encoders and proposed the fixed-slope universal lossy encoder for stationary ergodic sources.

In this chapter, we study the blockwise fixed-slope lossy coding of individual sequences, and formulate the optimal cost $\xi(x)$ for a given sequence $x \in \mathcal{X}^\infty$ attainable by blockwise lossy encoders. Let $\hat{S}(x)$ (resp. $\tilde{S}(x)$) be the optimal cost with respect to the overlapping (resp. non-overlapping) empirical distribution of the given sequence x. Our main result clarifies that three quantities $\xi(x)$, $\hat{S}(x)$, and $\tilde{S}(x)$ are equivalent. Moreover, we propose a fixed-slope universal lossy encoder based on the complexity function, and show that the optimal cost can be attained by this universal lossy encoder.

As an application of the result, we show that the blockwise fixed-slope lossy encoder based on the complexity function can attain asymptotically the optimal cost for any stationary ergodic source. We also investigate the blockwise fixed-slope lossy coding of nonstationary sources.

5.2 Fixed-Slope Lossy Coding of Individual Sequences

We consider the blockwise variable-rate lossy coding of individual sequences. Following the same line as in [33], we investigate the cost function associated with the variable-rate lossy block encoders. Fix a positive real number λ. For the lossy encoder (π_k, σ_k) of block size k ($k \in \mathcal{N}$), define the *cost* $s_k(a_1^k, \pi_k, \sigma_k)$ of each block $a_1^k \in \mathcal{X}$ as

$$s_k(a_1^k, \pi_k, \sigma_k) \triangleq \sigma_k\left(\pi_k(a_1^k)\right) + \lambda d_k\left(a_1^k, \pi_k(a_1^k)\right). \tag{5.1}$$

The multiplier λ is used to control the trade-off between the distortion and the compression rate. $-\lambda$ is interpreted as the slope of the line $R = -\lambda D + S$ in the D-R plain created by the various encoders (π_k, σ_k) such that

$$(D, R) = \left(\frac{1}{k} d_k(a_1^k, \pi_k(a_1^k)), \frac{1}{k}\sigma_k(\pi_k(a_1^k))\right)$$

and $S = s(a_1^k, \sigma_k, \pi_k)$.

For each sequence $x \in \mathcal{X}^\infty$, the asymptotic cost $\xi(x, \pi_k, \sigma_k)$ attained by the lossy encoder (π_k, σ_k) is defined as

$$\xi(x, \pi_k, \sigma_k) \triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{k} s_k(x_{tk+1}^{(t+1)k}, \pi_k, \sigma_k). \tag{5.2}$$

Furthermore, let $\xi(x)$ be the optimal cost for $x \in \mathcal{X}^\infty$ attainable by lossy block encoders defined as

$$\xi(x) \triangleq \inf_k \inf_{(\pi_k, \sigma_k)} \xi(x, \pi_k, \sigma_k),$$

where $\inf_{(\pi_k,\sigma_k)}$ is taken over all the lossy block encoders of block size k.

In order to clarify the optimal cost $\xi(x)$, we introduce two quantities $\tilde{S}(x)$ and $\hat{S}(x)$. Let $q_k(\cdot|x_1^{Tk})$ be the non-overlapping empirical distribution of $x_1^{Tk} \in \mathcal{X}^{Tk}$ defined in (2.3). Then, $\tilde{S}_k(x_1^{Tk})$ is defined as the optimal average cost with respect to $q_k(\cdot|x_1^{Tk})$:

$$\begin{aligned}\tilde{S}_k(x_1^{Tk}) &\stackrel{\triangle}{=} \inf_{(\pi_k,\sigma_k)} E_q\left[\frac{1}{k}s_k\left(X_1^k,\pi_k,\sigma_k\right)\right] \\ &= \inf_{(\pi_k,\sigma_k)} \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k),\end{aligned}$$

where inf is taken over all lossy encoders of block size k. Letting $T \to \infty$ and optimizing the block size k of the encoder, we define $\tilde{S}(x)$ as

$$\tilde{S}(x) \stackrel{\triangle}{=} \inf_k \limsup_{T\to\infty} \tilde{S}_k(x_1^{Tk}).$$

By the definitions, it is apparent that $\xi(x) \geq \tilde{S}(x)$ for any $x \in \mathcal{X}^\infty$.

Along with $\tilde{S}(x)$, we define $\hat{S}(x)$ using the overlapping empirical distribution $p_k(\cdot|x_1^n)$ defined in (2.1). Let $\hat{S}_k(x_1^n)$ be the optimal average cost with respect to $p_k(\cdot|x_1^n)$:

$$\begin{aligned}\hat{S}_k(x_1^n) &\stackrel{\triangle}{=} \inf_{(\pi_k,\sigma_k)} E_p\left[\frac{1}{k}s_k\left(X_1^k,\pi_k,\sigma_k\right)\right] \\ &= \inf_{(\pi_k,\sigma_k)} \frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}s_k(x_i^{i+k-1},\pi_k,\sigma_k).\end{aligned}$$

Then, $\hat{S}(x)$ is defined as

$$\hat{S}(x) \stackrel{\triangle}{=} \inf_k \limsup_{n\to\infty} \hat{S}_k(x_1^n).$$

Though one may suspect that the quantity $\hat{S}(x)$ has no implication, $\hat{S}(x)$ plays quite important roles in our discussions. As shown in Section 5.4, the overlapping version $\hat{S}(x)$ is a useful quantity for investigating the coding problem of the stochastic sources.

The next theorem shows that $\hat{S}(x)$ is equal to $\tilde{S}(x)$ and represents the optimal cost $\xi(x)$ attainable by lossy block encoders.

Theorem 5.1 For any $x \in \mathcal{X}^\infty$,

$$\xi(x) = \tilde{S}(x) = \hat{S}(x), \tag{5.3}$$

provided that there exists a reference letter r satisfying (3.2).

Next, we show that there exists a universal lossy block encoder which attains the optimal cost (5.3).

Fix the positive number λ and let L be an arbitrary complexity function. For each $k \in \mathcal{N}$, let π_k^L be the quantizer of block size k defined as

$$\pi_k^L(a_1^k) \triangleq \arg\min_{b_1^k \in \mathcal{Y}^k} \left[L(b_1^k) + \lambda d_k(a_1^k, b_1^k) \right]. \tag{5.4}$$

Since L satisfies the Kraft inequality (3.4), the pair (π_k^L, L) defines the lossy block encoder of block size k.

The next theorem shows that the lossy encoder (π_k^L, L) based on the complexity function L can attain the optimal cost.

Theorem 5.2 Let L be a complexity function. The sequence $\{(\pi_k^L, L)\}_{k=1}^{\infty}$ of the lossy block encoders satisfies that

$$\lim_{k\to\infty} \xi(x, \pi_k^L, L) = \tilde{S}(x) = \hat{S}(x) \quad \forall x \in \mathcal{X}^\infty, \tag{5.5}$$

provided that there exists a reference letter r satisfying (3.2).

Theorem 5.1 and Theorem 5.2 will be proved in the next section.

5.3 Proof of Theorems

In the proof of Theorem 5.1 and Theorem 5.2, we use the techniques developed in previous chapters. Recall that for $a_1^m \in \mathcal{X}^m$ and $r \in \mathcal{Y}$, $\Delta_k(a_1^m, r)$ is defined in (3.9) as

$$\Delta_k(a_1^m, r) \triangleq \sum_{i=1}^{k-1} d_1(a_i, r) + \sum_{i=m-k+2}^{m} d_1(a_i, r).$$

The following lemma plays an important role in the proofs.

Lemma 5.1 Let L be a complexity function and fix $r \in \mathcal{Y}$. For any $k \in \mathcal{N}$ and any $\gamma > 0$, there exists $m_0 = m_0(\gamma, k, \mathcal{Y}, L)$ such that for any (π_k, σ_k) and any $m \geq m_0$, the encoder (π_m^L, L) satisfies that

$$s_m(a_1^m, \pi_m^L, L) \leq \min_{0\leq j\leq k-1} \sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} s_k(a_{i+1}^{i+k}, \pi_k, \sigma_k) + m\gamma + \lambda\Delta_k(a_1^m, r) \quad \forall a_1^m \in \mathcal{X}^m,$$

$$
\begin{array}{cccccccc}
a_1 & \cdots & a_j & \underbrace{a_{j+1}^{j+k}} & \cdots & \underbrace{a_{\bar{j}-k+1}^{\bar{j}}} & a_{\bar{j}+1} \cdots & a_m \\
\downarrow & & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow \\
\boxed{\pi_1^*} & & \boxed{\pi_1^*} & \boxed{\pi_k} & & \boxed{\pi_k} & \boxed{\pi_1^*} & \boxed{\pi_1^*} \\
\downarrow & & \downarrow & \downarrow & & \downarrow & \downarrow & \downarrow
\end{array}
$$

Figure 5.1: Construction of $\pi_{k,m}^{(j)}$

and thus,

$$s_m(a_1^m, \pi_m^L, L) \leq \sum_{i=0}^{m-k} \frac{1}{k} s_k(a_{i+1}^{i+k}, \pi_k, \sigma_k) + m\gamma + \lambda \Delta_k(a_1^m, r) \quad \forall a_1^m \in \mathcal{X}^m.$$

Proof: For a given quantizer π_k and each j $(0 \leq j \leq k-1)$, let $\pi_{k,m}^{(j)}$ be the quantizer which quantizes a block $a_1^m \in \mathcal{X}^m$ in the following manner. At first, the first j symbols $a_1, \cdots, a_j$ of a_1^m are quantized by the quantizer π_1^* such that $\pi_1^*(a) = r$ for all $a \in \mathcal{X}$. Then, the following a_{j+1}^m is partitioned into blocks of size k and each block is quantized by π_k. Lastly, the remaining symbols $a_{\bar{j}+1}, \cdots, a_m$ $(m - \bar{j} < k)$ of a_1^m are quantized by π_1^* (see Fig. 5.1). Since d_k is subadditive, it is apparent that for any π_k and j, $\pi_{k,m}^{(j)}$ satisfies that

$$
\begin{aligned}
d_m\left(a_1^m, \pi_{k,m}^{(j)}(a_1^m)\right) &\leq \sum_{i=1}^{j} d_1(a_i, \pi_1^*(a_i)) \\
&\quad + \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) \\
&\quad + \sum_{i=\bar{j}+1}^{m} d_1(a_i, \pi_1^*(a_i)) \\
&\leq \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) + \Delta_k(a_1^m, r). \qquad (5.6)
\end{aligned}
$$

On the other hand, by Lemma 3.1, for any $\gamma > 0$, there exists $m_0 = m_0(\gamma, k, \mathcal{Y}, L)$ such that for any $m \geq m_0$ and $b_1^m \equiv \pi_{k,m}^{(j)}(a_1^m)$ $(0 \leq j \leq k-1)$,

$$
\begin{aligned}
L(b_1^m) &\leq \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} \sigma_k\left(b_{i+1}^{i+k}\right) + m\gamma \\
&= \sum_{\substack{i \equiv j \bmod k \\ 0 \leq i \leq m-k}} \sigma_k\left(\pi_k(a_{i+1}^{i+k})\right) + m\gamma. \qquad (5.7)
\end{aligned}
$$

Thus, for any k, any (π_k, σ_k), any j $(0 \leq j \leq k-1)$, and any $a_1^m \in \mathcal{X}^m$, we have

$$\begin{aligned}
s_m(a_1^m, \pi_m^L, L) &\leq s_m(a_1^m, \pi_{k,m}^{(j)}, L) && (5.8)\\
&= L\left(\pi_k^{(j)}(a_1^m)\right) + \lambda d_m\left(a_1^m, \pi_{k,m}^{(j)}(a_1^m)\right)\\
&\leq L\left(\pi_k^{(j)}(a_1^m)\right)\\
&\quad +\lambda\left[\sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) + \Delta_k(a_1^m, r)\right] && (5.9)\\
&\leq \left[\sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} \sigma_k\left(\pi_k(a_{i+1}^{i+k})\right) + m\gamma\right]\\
&\quad +\lambda\left[\sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} d_k(a_{i+1}^{i+k}, \pi_k(a_{i+1}^{i+k})) + \Delta_k(a_1^m, r)\right] && (5.10)\\
&= \sum_{\substack{i\equiv j \bmod k\\ 0\leq i\leq m-k}} s_k(a_{i+1}^{i+k}, \pi_k, \sigma_k) + m\gamma + \lambda\Delta_k(a_1^m, r), && (5.11)
\end{aligned}$$

where the inequality (5.8) follows from (5.4), the inequality (5.9) from (5.6), and the inequality (5.10) from (5.7). □

Lemma 5.2 If $x \in \mathcal{X}^\infty$ and $r \in \mathcal{Y}$ satisfy (3.2) then

$$\limsup_{k\to\infty} \xi(x, \pi_k^L, L) \leq \hat{S}(x).$$

Proof: Fix $\gamma > 0$, and choose k and n_0 such that for any $n \geq n_0$,

$$\inf_{(\pi_k,\sigma_k)} \frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}s_k(x_i^{i+k-1}, \pi_k, \sigma_k) \leq \hat{S}(x) + \gamma.$$

For each $n \geq n_0$, we can choose an encoder (π_k, σ_k) such that

$$\frac{1}{n-k+1}\sum_{i=1}^{n-k+1}\frac{1}{k}s_k(x_i^{i+k-1}, \pi_k, \sigma_k) \leq \hat{S}(x) + 2\gamma, \qquad (5.12)$$

where (π_k, σ_k) may depend on n. Since (3.2) holds, by Lemma 5.1, there exists $m_0 = m_0(\gamma, k, \mathcal{Y}, L)$ such that for any $m \geq m_0$, $n \geq n_0$, and t $((t+1)m \leq n)$, we have

$$s_m(x_{tm+1}^{(t+1)m}, \pi_m^L, L)$$

$$
\begin{aligned}
&\leq \sum_{i=0}^{m-k} \frac{1}{k} s_k(x_{tm+i+1}^{tm+i+k}, \pi_k, \sigma_k) + m\gamma + \lambda \Delta_k(x_{tm+1}^{(t+1)m}, r), \\
&= \sum_{i=tm+1}^{(t+1)m-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) + m\gamma + \lambda \Delta_k(x_{tm+1}^{(t+1)m}, r).
\end{aligned}
$$

It should be noted that m_0 depends on neither (π_k, σ_k) nor n. Hence, we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{m} s_m(x_{tm+1}^{(t+1)m}, \pi_m^L, L) \\
&\quad\leq \frac{1}{Tm}\sum_{t=0}^{T-1} \sum_{i=tm+1}^{(t+1)m-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) + \gamma + \frac{\lambda}{Tm}\sum_{t=0}^{T-1} \Delta_k(x_{tm+1}^{(t+1)m}, r) \\
&\quad\leq \frac{1}{Tm}\sum_{i=1}^{Tm-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) + \gamma + \frac{\lambda}{Tm}\sum_{t=0}^{T-1} \Delta_k(x_{tm+1}^{(t+1)m}, r) \\
&\quad\leq \frac{1}{Tm-k+1}\sum_{i=1}^{Tm-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) \\
&\quad\quad + \gamma + \frac{\lambda}{Tm}\sum_{t=0}^{T-1} \Delta_k(x_{tm+1}^{(t+1)m}, r). \qquad (5.13)
\end{aligned}
$$

As a same manner as (3.20), we can show that for sufficiently large T and m,

$$
\frac{\lambda}{Tm}\sum_{t=0}^{T-1} \Delta_k(x_{tm+1}^{(t+1)m}, r) \leq \gamma. \qquad (5.14)
$$

Combining (5.12), (5.13), and (5.14), for sufficiently large T and m,

$$
\frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{m} s_m(x_{tm+1}^{(t+1)m}, \pi_m^L, L) \leq \hat{S}(x) + 4\gamma.
$$

Since $\gamma > 0$ can be chosen arbitrarily small, we have

$$
\limsup_{m\to\infty} \limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \frac{1}{m} s_m(x_{tm+1}^{(t+1)m}, \pi_m^L, L) \leq \hat{S}(x).
$$

□

Lemma 5.3 If $x \in \mathcal{X}^\infty$ and $r \in \mathcal{Y}$ satisfy (3.2) then

$$
\hat{S}(x) \leq \tilde{S}(x).
$$

Proof: Notice that

$$\begin{aligned}
&\hat{S}(x)\\
&= \inf_m \limsup_{n\to\infty} \inf_{(\pi_m,\sigma_m)} \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m,\sigma_m)\\
&\le \inf_m \inf_{(\pi_m,\sigma_m)} \limsup_{n\to\infty} \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m,\sigma_m)\\
&\le \limsup_{m\to\infty}\limsup_{n\to\infty} \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L).
\end{aligned}$$

Hence, to show the theorem, it is sufficient to show that

$$\limsup_{m\to\infty}\limsup_{n\to\infty} \frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L) \le \tilde{S}(x). \tag{5.15}$$

Fix $\gamma > 0$, and choose k and T_0 such that for any $T \ge T_0$,

$$\inf_{(\pi_k,\sigma_k)} \frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k) \le \tilde{S}(x)+\gamma.$$

For each T ($T \ge T_0$), we can choose the encoder (π_k,σ_k) such that

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{1}{k}s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k) \le \tilde{S}(x)+2\gamma, \tag{5.16}$$

where (π_k,σ_k) may depend on T. By Lemma 5.1, there exists $m_0 = m_0(\gamma,k,\mathcal{Y},L)$ such that for any $m \ge m_0$, $T \ge T_0$, and i ($i \le Tk-m+1$), we have

$$s_m(x_i^{i+m-1},\pi_m^L,L) \le \min_{0\le j\le k-1} \sum_{\substack{l\equiv j \bmod k\\ 0\le l\le m-k}} s_k(x_{l+i}^{l+k+i-1},\pi_k,\sigma_k)+m\gamma+\lambda\Delta_k(x_i^{i+m-1},r).$$

Especially, we have

$$\begin{aligned}
s_m(x_i^{i+m-1},\pi_m^L,L) &\le \sum_{\substack{l\equiv j_i \bmod k\\ 0\le l\le m-k}} s_k(x_{l+i}^{l+k+i-1},\pi_k,\sigma_k)+m\gamma+\lambda\Delta_k(x_i^{i+m-1},r),\\
&= \sum_{\substack{t\\ i\le (tk+1)\le i+m-k}} s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k)+m\gamma+\lambda\Delta_k(x_i^{i+m-1},r),
\end{aligned}$$

where j_i is chosen so that $i+j_i \equiv 1 \bmod k$. Hence, we have

$$\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L)$$

$$\begin{aligned}
&\leq \frac{1}{m}\sum_{i=1}^{n-m+1}\sum_{\substack{t\\ i\leq(tk+1)\leq i+m-k}} s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k)+(n-m+1)\gamma \\
&\quad +\frac{\lambda}{m}\sum_{i=1}^{n-m+1}\Delta_k(x_i^{i+m-1},r) \\
&\leq \sum_{t=0}^{\lfloor n/k\rfloor-1} s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k)+(n-m+1)\gamma+\frac{2\lambda k}{m}\sum_{i=1}^{n} d_1(x_i,r), \qquad (5.17)
\end{aligned}$$

where (5.17) follows from that each k-block is encoded by (π_k,σ_k) at most m times, and each symbol is quantized by π_1^* at most $2k$ times. Since (5.17) holds for any n and T $(n\geq T)$, letting $T=\lfloor n/k\rfloor$, we have

$$\begin{aligned}
&\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L) \\
&\quad\leq \frac{1}{n-m+1}\sum_{t=0}^{T-1} s_k(x_{tk+1}^{(t+1)k},\pi_k,\sigma_k) \\
&\quad\quad +\gamma+\frac{2\lambda k}{m(n-m+1)}\sum_{i=1}^{n} d_1(x_i,r). \qquad (5.18)
\end{aligned}$$

Since (5.16) holds for $T\geq T_0$, the first term of the right hand side of (5.18) is bounded by

$$\frac{1}{n-m+1}\sum_{t=0}^{T-1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L)\leq\left(1+\frac{m}{Tk-m}\right)\left(\tilde{S}(x)+2\gamma\right). \qquad (5.19)$$

On the other hand, as a same manner as (3.17), we can show that for sufficiently large n and m,

$$\frac{2\lambda k}{m(n-m+1)}\sum_{i=1}^{n} d_1(x_i,r)\leq\gamma. \qquad (5.20)$$

Thus, if n is so large that $T\geq T_0$, then, combining (5.18), (5.19), and (5.20), we have

$$\begin{aligned}
&\frac{1}{n-m+1}\sum_{i=1}^{n-m+1}\frac{1}{m}s_m(x_i^{i+m-1},\pi_m^L,L) \\
&\quad\leq \left(1+\frac{m}{Tk-m}\right)\left(\tilde{S}(x)+2\gamma\right)+2\gamma \\
&\quad\leq \left(1+\frac{m}{n-k-m}\right)\left(\tilde{S}(x)+2\gamma\right)+2\gamma.
\end{aligned}$$

Since $m/(n-k-m) \to 0$ as $n \to \infty$, we have, for sufficiently large n,

$$\frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \frac{1}{m} s_m(x_i^{i+m-1}, \pi_m^L, L) \le \tilde{S}(x) + 5\gamma.$$

Since $\gamma > 0$ can be chosen arbitrarily small, we have (5.15). $\square$

Proof of Theorem 5.1 and Theorem 5.2: By the definitions, it is apparent that

$$\tilde{S}(x) \le \xi(x) \le \liminf_{k\to\infty} \xi(x, \pi_k^L, L),$$

for any $x \in \mathcal{X}^\infty$. Moreover, by Lemma 5.2 and Lemma 5.3, we have

$$\limsup_{k\to\infty} \xi(x, \pi_k^L, L) \le \hat{S}(x) \le \tilde{S}(x) \quad \forall x \in \mathcal{X}^\infty,$$

provided that there exists a reference letter r satisfying (3.2). Hence, we have

$$\lim_{k\to\infty} \xi(x, \pi_k^L, L) = \xi(x) = \tilde{S}(x) = \hat{S}(x) \quad \forall x \in \mathcal{X}^\infty,$$

provided that there exists a reference letter r satisfying (3.2). $\square$

5.4 Fixed-Slope Lossy Coding of Nonstationary Sources

In this section, we apply our result on the fixed-slope lossy coding of individual sequences to investigate the fixed-slope lossy coding of stochastic sources.

In Subsection 5.4.1, we study the blockwise fixed-slope lossy coding of stationary ergodic sources. While Yang *et al.* [33] showed a similar result, their coding scheme requires that the given whole sequence x_1^n is quantized into y_1^n to attain the optimal cost. On the other hand, our result ensures that we can attain the optimal cost even if we fix the block size k of the encoder and encode each blocks $x_{tk+1}^{(t+1)k}$ ($t = 0, 1, \cdots$) separately, provided that the block size k of the encoder is sufficiently large.

In Subsection 5.4.2, we consider the fixed-slope lossy coding of nonstationary sources. We generalize the result of Yang *et al.* [33] and clarify the optimal cost for nonstationary sources with finite alphabet.

5.4.1 Fixed-Slope Lossy Coding of Ergodic Sources

In this subsection, we assume that $\mathcal{X}$ is a standard alphabet and that the distortion measure $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, \infty)$ is a measurable function. In addition, we introduce some assumptions appeared in [33]. We assume that the distortion measure is additive, and that there exists a reference letter $r \in \mathcal{Y}$ satisfying (3.25). Moreover, we assume that

$$\sup_{a\in\mathcal{X}} \min_{b\in\mathcal{Y}} d_1(a, b) = 0, \tag{5.21}$$

otherwise, we may replace d_1 by

$$d_1'(a, b) \equiv d_1(a, b) - \min_{b\in\mathcal{Y}} d_1(a, b).$$

Let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ be a stationary ergodic source. Under conditions (3.25) and (5.21), the argument D in the rate-distortion function $R(D|\boldsymbol{X})$ and the argument R in the distortion-rate function $D(R|\boldsymbol{X})$ can vary from 0 to $+\infty$, respectively. Since the rate-distortion function $R(D|\boldsymbol{X})$ is a convex function of D, $R(D|\boldsymbol{X})$ is continuous and the right-hand derivative $R_+'(D|\boldsymbol{X})$ at D exists for $D \in (0, D_{\max})$, where $D_{\max} \stackrel{\triangle}{=} \inf\{D : R(D|\boldsymbol{X}) = 0\}$. For any $\lambda > 0$, define $D_\lambda(\boldsymbol{X}) \stackrel{\triangle}{=} \inf\{D : D \geq 0, R_+'(D|\boldsymbol{X}) > -\lambda\}$ and $R_\lambda(\boldsymbol{X}) \stackrel{\triangle}{=} R(D_\lambda(\boldsymbol{X})|\boldsymbol{X})$. Then, define $S(\boldsymbol{X})$ as $S(\boldsymbol{X}) \stackrel{\triangle}{=} R_\lambda(\boldsymbol{X}) + \lambda D_\lambda(\boldsymbol{X})$.

Yang *et al.* established the following results.

Theorem 5.3 ([33]) For any stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$, the following two statements hold.

1. (converse part)

   For any sequence $\{(\pi_n, \sigma_n)\}_{n=1}^{\infty}$ of variable-rate lossy encoders,

   $$\liminf_{n\to\infty} \frac{1}{n} s_n(x_1^n, \pi_n, \sigma_n) \geq S(\boldsymbol{X}), \tag{5.22}$$

   for almost all $x \in \mathcal{X}^\infty$.

2. (direct part)

   There exists a sequence $\{(\pi_n, \sigma_n)\}_{n=1}^{\infty}$ of variable-rate encoders such that

   $$\limsup_{n\to\infty} \frac{1}{n} s_n(x_1^n, \pi_n, \sigma_n) \leq S(\boldsymbol{X}), \tag{5.23}$$

   for almost all $x \in \mathcal{X}^\infty$.

Moreover, they showed that a universal lossless encoder can be used to construct a fixed-slope universal lossy encoder.

Theorem 5.4 ([33]) Suppose that $\{(\pi_n^*, \sigma_n^*)\}_{n=1}^\infty$ be a sequence of variable-rate lossy encoders satisfying that

$$s_n(a_1^k, \pi_n^*, \sigma_n^*) = \min_{b_1^n \in \mathcal{Y}^n} [\sigma_n^*(b_1^n) + \lambda d_n(a_1^n, b_1^n)], \quad \forall n \in \mathcal{N}, \forall a_1^n \in \mathcal{X}^n. \quad (5.24)$$

Further, assume that for any stationary ergodic source $\boldsymbol{Y} = \{Y_i\}_{i=1}^\infty$ with alphabet $\mathcal{Y}$,

$$\limsup_{n\to\infty} \frac{1}{n}\sigma_n^*(y_1^n) \leq H(\boldsymbol{Y}), \quad (5.25)$$

for almost all $y \in \mathcal{Y}^\infty$. Then, $\{(\pi_n^*, \sigma_n^*)\}_{n=1}^\infty$ satisfies that for any stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$,

$$\limsup_{n\to\infty} \frac{1}{n} s_n(x_1^n, \pi_n^*, \sigma_n^*) \leq S(\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$.

Now, we show that the blockwise lossy encoder (π_k^L, L) based on the complexity function L is optimal for stationary ergodic sources.

Lemma 5.4 Let L be arbitrary complexity function and π_k^L be the quantizer defined in (5.4). Then, the sequence $\{(\pi_k^L, L)\}_{i=1}^\infty$ of variable-rate lossy encoders satisfies that for any stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$,

$$\limsup_{n\to\infty} \frac{1}{n} s_n(x_1^n, \pi_n^L, L) \leq S(\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$.

Proof: Since (5.4) implies (5.24), to show the lemma, it is sufficient to show that

$$\limsup_{n\to\infty} \frac{1}{n} L(y_1^n) \leq H(\boldsymbol{Y}), \quad (5.26)$$

for almost all $y \in \mathcal{Y}^\infty$.

Fix $\epsilon > 0$. By Theorem 2.2, for sufficiently large m and T, there exists a length function σ_m^T on $\mathcal{Y}^m$ such that

$$\frac{1}{T}\sum_{i=0}^{T-1} \frac{1}{m}\sigma_m^T(y_{im+1}^{(i+1)m}) \leq \hat{H}(y) + \epsilon/2.$$

Moreover, by Lemma 3.1, for sufficiently large n, letting $T = \lfloor n/m \rfloor$, we have

$$L(y_1^n) \le \sum_{\substack{0 \le i \le n-m \\ i \equiv 0 \bmod m}} \sigma_m^T(y_{i+1}^{i+m}) + n\epsilon/2,$$

and thus, for sufficiently large n,

$$\frac{1}{n} L(y_1^n) \le \hat{H}(y) + \epsilon.$$

Since we can take $\epsilon > 0$ arbitrarily small, we have

$$\limsup_{n\to\infty} \frac{1}{n} L(y_1^n) \le \hat{H}(y) \quad \forall y \in \mathcal{Y}^\infty.$$

On the other hand, it is known that $H(y) = H(\boldsymbol{Y})$ for almost all $y \in \mathcal{Y}^\infty$ [36]. Hence, we have (5.26). □

Next, we show that the optimal cost can be attained even if we encode each blocks $x_{tk+1}^{(t+1)k}$ $(t = 0, 1, \cdots)$ separately.

Theorem 5.5 Let L be a complexity function. Then, for any stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$, we have

$$\lim_{k\to\infty} \xi(x, \pi_k^L, L) = S(\boldsymbol{X}), \tag{5.27}$$

for almost all $x \in \mathcal{X}^\infty$.

Proof: The converse part of Theorem 5.3 implies that

$$\lim_{k\to\infty} \xi(X, \pi_k^L, L) \ge S(\boldsymbol{X}),$$

for almost all $x \in \mathcal{X}^\infty$. Thus, to show the theorem, it is sufficient to show that for almost all $x \in \mathcal{X}^\infty$,

$$\lim_{k\to\infty} \xi(X, \pi_k^L, L) \le S(\boldsymbol{X}).$$

By the assumption (3.25) and Lemma 3.5, r satisfies (3.2) for almost all $x \in \mathcal{X}^\infty$, and thus, we can apply Theorem 5.2 to almost all x. Hence, we have

$$\begin{aligned} &\lim_{k\to\infty} \xi(x, \pi_k^L, L) \\ &= \hat{S}(X) \end{aligned} \tag{5.28}$$

$$
\begin{aligned}
&= \inf_k \limsup_{n\to\infty} \inf_{(\pi_k,\sigma_k)} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) && (5.29)\\
&\le \inf_k \inf_{(\pi_k,\sigma_k)} \limsup_{n\to\infty} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \frac{1}{k} s_k(x_i^{i+k-1}, \pi_k, \sigma_k) \\
&= \inf_k \inf_{(\pi_k,\sigma_k)} E\left[\frac{1}{k} s_k(X_1^k, \pi_k, \sigma_k)\right] && (5.30)\\
&= S(\boldsymbol{X}), && (5.31)
\end{aligned}
$$

where (5.28) follows from Theorem 5.2, (5.29) from the definition of $\hat{S}$, (5.30) from the ergodic theorem [5], and (5.31) from the direct part of Theorem 5.3.

□

5.4.2 Fixed-Slope Lossy Coding of Nonstationary Sources

In this subsection, we assume that $\mathcal{X}$ is finite, and that the distortion measure is additive and satisfies (5.21).

At first, we consider the fixed-slope lossy coding problem of stationary sources. Let $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ be a stationary source. By the ergodic decomposition theorem [5], there exists a measure $\phi_{\boldsymbol{X}}$ on the sources such that

$$\phi_{\boldsymbol{X}}\Big(\{\boldsymbol{V} : \boldsymbol{V} = \{V_i\}_{i=1}^\infty \text{ is stationary ergodic}\}\Big) = 1,$$

and

$$P_{X_1^\infty}(A) = \int P_{V_1^\infty}(A) d\phi_{\boldsymbol{X}}(\boldsymbol{V}),$$

for all measurable set $A \subset \mathcal{X}^\infty$. For a stationary source $\boldsymbol{X}$, let $\bar{S}(\boldsymbol{X})$ be the average of the optimal cost $S(\boldsymbol{V})$ of the ergodic component $\boldsymbol{V}$ of $\boldsymbol{X}$, that is,

$$\bar{S}(\boldsymbol{X}) \triangleq \int S(\boldsymbol{V}) d\phi_{\boldsymbol{X}}(\boldsymbol{V}).$$

The next theorem clarifies that the optimal average cost for a stationary source $\boldsymbol{X}$ equals to $\bar{S}(\boldsymbol{X})$.

Theorem 5.6 The following two statements hold.

1. (converse part)

   For any stationary source $\boldsymbol{X} = \{X_i\}_{i=1}^\infty$ and any sequence $\{(\pi_n, \sigma_n)\}_{n=1}^\infty$ of encoders,
   $$\liminf_{n\to\infty} \frac{1}{n} E\left[s_n(X_1^n, \pi_n, \sigma_n)\right] \ge \bar{S}(\boldsymbol{X}). \qquad (5.32)$$

2. (direct part)

   Let L be arbitrary complexity function and π_k^L be the quantizer defined in (5.4). Then, for any stationary source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$,

$$\limsup_{n\to\infty} \frac{1}{n} E\left[s_n(X_1^n, \pi_n^L, L)\right] \le \bar{S}(\boldsymbol{X}). \tag{5.33}$$

Before proving Theorem 5.6, we show that the cost associated with the encoder (π_k^L, L) is bounded.

Lemma 5.5 For the encoders $\{(\pi_k^L, L)\}$ based on the complexity function L, there exists $M = M(L) < \infty$ such that

$$s_k(a_1^k, \pi_k^L, L) \le kM \quad \forall k \in \mathcal{N}, a_1^k \in \mathcal{X}^k. \tag{5.34}$$

Proof: Let $M_0 = \lceil \log |\mathcal{Y}| \rceil$ and fix $\bar{b} \in \mathcal{Y}$ arbitrarily. For each $k \in \mathcal{N}$, let $\bar{b}_1^k \in \mathcal{Y}^k$ be the sequence such that $\bar{b}_i = \bar{b}$ $(i = 1, \cdots, k)$. It is not hard to see that there exists k_0 such that for any $k \ge k_0$,

$$L(\bar{b}_1^k) \le kM_0.$$

For each $k = 1, \cdots, k_0 - 1$, let $M_k \triangleq (1/k)L(\bar{b}_1^k)$, and then, let M be

$$M = \max_{0\le k\le k_0-1} M_k + \lambda d_{\max},$$

where $d_{\max} \triangleq \max_{a\in\mathcal{X}} \max_{b\in\mathcal{Y}} d_1(a,b) < \infty$. By the definitions, for each $k \in \mathcal{N}$, $\bar{b}_1^k \in \mathcal{Y}^k$ satisfies that

$$L(\bar{b}_1^k) + \lambda d_k(a_1^k, \bar{b}_1^k) \le kM \quad \forall a_1^k \in \mathcal{X}^k.$$

Thus, by the definition of π_k^L, we have (5.34). □

Proof of Theorem 5.6: At first, we prove (5.32). Fix $\epsilon > 0$ and $\{(\pi_n, \sigma_n)\}_{n=1}^{\infty}$. By (5.22), for any stationary and ergodic source $\boldsymbol{V} = \{V_i\}_{i=1}^{\infty}$ with the alphabet $\mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{n} s_n(V_1^n, \pi_n, \sigma_n) \ge S(\boldsymbol{V}) \quad \text{a.s.}$$

Thus, there exists $n_0 = n_0(\epsilon, \boldsymbol{X})$ and a set Γ_1 of ergodic components of $\boldsymbol{X}$ such that

$$\phi_{\boldsymbol{X}}(\Gamma_1) \ge 1 - \epsilon, \tag{5.35}$$

and for any $\boldsymbol{V} \in \Gamma_1$ and any $n \geq n_0$,

$$\Pr\left\{\frac{1}{n}s_n(V_1^n, \pi_n, \sigma_n) \geq S(\boldsymbol{V}) - \epsilon\right\} \geq 1 - \epsilon.$$

For any $\boldsymbol{V} \in \Gamma_1$ and any $n \geq n_0$,

$$E\left[\frac{1}{n}s_n(V_1^n, \pi_n, \sigma_n)\right] \geq (1-\epsilon)\left(S(\boldsymbol{V}) - \epsilon\right). \tag{5.36}$$

By the ergodic decomposition theorem [5] combined with (5.35) and (5.36), we have for any $n \geq n_0$,

$$\begin{aligned} E\left[\frac{1}{n}s_k(X_1^n, \pi_n, \sigma_n)\right] &= \int E\left[\frac{1}{n}s_n(V_1^n, \pi_n, \sigma_n)\right] d\phi_{\boldsymbol{X}}(\boldsymbol{V}) \\ &\geq (1-\epsilon)(\bar{S}(\boldsymbol{X}) - \epsilon) - \epsilon\left[\sup_{\boldsymbol{V}} S(\boldsymbol{V})\right]. \end{aligned}$$

Since $S(\boldsymbol{V}) \leq \lceil \log|\mathcal{Y}| \rceil + \lambda d_{\max} < \infty$ for any $\boldsymbol{V}$, and $\epsilon > 0$ can be chosen arbitrarily small, we have, for any $\epsilon' > 0$ and sufficiently large n,

$$E\left[\frac{1}{n}s_n(X_1^n, \pi_n, \sigma_n)\right] \geq \bar{S}(\boldsymbol{X}) - \epsilon'.$$

Thus, we have (5.32).

Next, we prove (5.33). Fix $\epsilon > 0$. Lemma 5.4 and Lemma 5.4 guarantees that $\{(\pi_n^L, L)\}_{n=1}^{\infty}$ satisfies that for any stationary ergodic source $\boldsymbol{V} = \{V_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{n}s_n(v_1^n, \pi_n^L, L) \leq S(\boldsymbol{V}),$$

for almost all $v \in \mathcal{X}^\infty$. Hence, for any stationary source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$, there exists $n_0 = n_0(\epsilon, \boldsymbol{X})$ and a set Γ_2 of ergodic components of $\boldsymbol{X}$ such that

$$\phi_{\boldsymbol{X}}(\Gamma_2) \geq 1 - \epsilon, \tag{5.37}$$

and for any $\boldsymbol{V} \in \Gamma_2$ and any $n \geq n_0$,

$$\Pr\left\{\frac{1}{n}s_n(V_1^n, \pi_n^L, L) \leq S(\boldsymbol{V}) + \epsilon\right\} \geq 1 - \epsilon. \tag{5.38}$$

By (5.34) and (5.38), for any $\boldsymbol{V} \in \Gamma_2$ and any $n \geq n_0$,

$$E\left[\frac{1}{n}s_n(V_1^n, \pi_n^L, L)\right] \leq S(V) + \epsilon + \epsilon M. \tag{5.39}$$

By the ergodic decomposition theorem combined with (5.34), (5.37), and (5.39), we have for any $n \geq n_0$,

$$
\begin{aligned}
E\left[\frac{1}{n}s_n(X_1^n,\pi_n^L,L)\right] &= \int E\left[\frac{1}{n}s_n(V_1^n,\pi_n^L,L)\right]d\phi_{\boldsymbol{X}}(\boldsymbol{V}) \\
&\leq \bar{S}(\boldsymbol{X})+\epsilon+\epsilon M+\epsilon\sup_{\boldsymbol{V}} E\left[\frac{1}{n}s_n(V_1^n,\pi_n^L,L)\right] \\
&\leq \bar{S}(\boldsymbol{X})+\epsilon+\epsilon M+\epsilon M.
\end{aligned}
$$

Since we can choose $\epsilon > 0$ arbitrarily small, we have (5.33). □

Next, we consider the blockwise lossy coding of nonstationary sources. To clarify the optimal cost, we introduce the *stationary hull* $\Lambda(\boldsymbol{X})$ of the source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ [11]. The stationary hull $\Lambda(\boldsymbol{X})$ of $\boldsymbol{X}$ is defined as the set of all sources $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X}$ such that for some sequence of integers $n_1 < n_2 < \cdots$,

$$
\lim_{j\to\infty}\frac{1}{n_j}\sum_{i=1}^{n_j} E\left[f(X_i,X_{i+1},\cdots)\right] = E\left[f(Z_1,Z_2,\cdots)\right],
$$

for every real-valued function f defined for sequences from $\mathcal{X}$ which depends on only finitely many coordinates. It is known that every $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ in $\Lambda(\boldsymbol{X})$ is stationary [11]. Let

$$
S^*(\boldsymbol{X}) \stackrel{\triangle}{=} \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} \bar{S}(\boldsymbol{Z}).
$$

The next theorem clarify that the optimal average cost for a source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ attainable by blockwise encoders is equal to $S^*(\boldsymbol{X})$.

Theorem 5.7 The following two statements hold.

1. (converse part)

   For any source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ and any lossy block encoder (π_k, σ_k),

$$
\limsup_{T\to\infty}\frac{1}{Tk}\sum_{t=0}^{T-1} E\left[s_k(X_{tk+1}^{(t+1)k},\pi_k,\sigma_k)\right] \geq S^*(\boldsymbol{X}). \tag{5.40}
$$

2. (direct part)

   Let L be arbitrary complexity function and π_k^L be the quantizer defined in (5.4). Then, for any source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$,

$$
\lim_{k\to\infty}\limsup_{T\to\infty}\frac{1}{Tk}\sum_{t=0}^{T-1} E\left[s_k(X_{tk+1}^{(t+1)k},\pi_k^L,L)\right] \leq S^*(\boldsymbol{X}). \tag{5.41}
$$

Before proving Theorem 5.7, we state some lemmas.

Lemma 5.6 For any stationary source $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$,

$$\inf_k \inf_{(\pi_k,\sigma_k)} E\left[\frac{1}{k} s_k(Z_1^k, \pi_k, \sigma_k)\right] = \liminf_{k\to\infty} \inf_{(\pi_k,\sigma_k)} E\left[\frac{1}{k} s_k(Z_1^k, \pi_k, \sigma_k)\right].$$

Proof: For any k_1, k_2, k_3 ($k_1 = k_2 + k_3$) and any σ_{k_2} and σ_{k_3}, there exists σ_{k_1} such that

$$\sigma_{k_1}(b_1^{k_1}) = \sigma_{k_2}(b_1^{k_2}) + \sigma_{k_3}(b_{k_2+1}^{k_2+k_3}) \quad \forall b_1^{k_1} \in \mathcal{Y}^{k_1}.$$

Since d_k is additive and $\boldsymbol{Z}$ is stationary, we have,

$$\begin{aligned} &\inf_{(\pi_{k_1},\sigma_{k_1})} E\left[s_{k_1}(Z_1^{k_1}, \pi_{k_1}, \sigma_{k_1})\right] \\ &\le \inf_{(\pi_{k_2},\sigma_{k_2})} E\left[s_{k_2}(Z_1^{k_2}, \pi_{k_2}, \sigma_{k_2})\right] + \inf_{(\pi_{k_3},\sigma_{k_3})} E\left[s_{k_3}(Z_1^{k_3}, \pi_{k_3}, \sigma_{k_3})\right]. \end{aligned}$$

Thus, we have the lemma. □

Lemma 5.7 For any $k \in \mathcal{N}$ and any $\gamma > 0$, there exists $n_0 = n_0(\gamma, k, \mathcal{Y})$ such that for any $n \ge n_0$ and any stationary source $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$,

$$E\left[\frac{1}{n} s_n(Z_1^n, \pi_n^L, L)\right] \le E\left[\frac{1}{k} s_k(Z_1^k, \pi_k^L, L)\right] + \gamma.$$

Proof: By Lemma 5.1, for any $k \in \mathcal{N}$ and $\gamma > 0$, there exists $n_0 = n_0(\gamma, k, \mathcal{Y})$ such that for any $n \ge n_0$ and any $a_1^n \in \mathcal{X}^n$,

$$s_n(a_1^n, \pi_n^L, L) \le \sum_{\substack{i \equiv 0 \bmod k \\ 0 \le i \le n-k}} s_k(a_{i+1}^{i+k}, \pi_k^L, L) + n\gamma.$$

Since $\boldsymbol{Z}$ is stationary, taking the expectation of both sides of the inequality, we have the lemma. □

For any source $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ and $\hat{\boldsymbol{Z}} = \{\hat{Z}_i\}_{i=1}^{\infty}$, let $\nu(\boldsymbol{Z}, \hat{\boldsymbol{Z}})$ be the distance between $\boldsymbol{Z}$ and $\hat{\boldsymbol{Z}}$ defined as

$$\nu(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) \triangleq \sum_{k=1}^{\infty} 2^{-k} \left(\sum_{a_1^k \in \mathcal{X}^k} |P_{Z_1^k}(a_1^k) - P_{\hat{Z}_1^k}(a_1^k)| \right).$$

The next theorem shows the average cost associated with (π_k^L, L) is continuous.

Lemma 5.8 For any $k \in \mathcal{N}$ and any $\gamma > 0$, there exists $\zeta = \zeta(k, \gamma) > 0$ such that if $\nu(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) < \zeta$ then

$$\left| E\left[\frac{1}{k} s_k(Z_1^k, \pi_k^L, L)\right] - E\left[\frac{1}{k} s_k(\hat{Z}_1^k, \pi_k^L, L)\right] \right| \leq \gamma.$$

Proof: Let M be the number satisfying (5.34), and fix $\epsilon > 0$ such that $\epsilon M \leq \gamma$. Choose ζ so that $\zeta \leq \epsilon/2^k$. Then, if $\nu(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) < \zeta$ then

$$\sum_{a_1^k \in \mathcal{X}^k} |P_{Z_1^k}(a_1^k) - P_{\hat{Z}_1^k}(a_1^k)| \leq \epsilon.$$

Thus,

$$\begin{aligned}
&\left| E\left[\frac{1}{k} s_k(Z_1^k, \pi_k^L, L)\right] - E\left[\frac{1}{k} s_k(\hat{Z}_1^k, \pi_k^L, L)\right] \right| \\
&= \left| \sum_{a_1^k \in \mathcal{X}^k} \frac{1}{k} s_k(a_1^k, \pi_k^L, L)[P_{Z_1^k}(a_1^k) - P_{\hat{Z}_1^k}(a_1^k)] \right| \\
&\leq M \sum_{a_1^k \in \mathcal{X}^k} |P_{Z_1^k}(a_1^k) - P_{\hat{Z}_1^k}(a_1^k)| \\
&\leq \epsilon M \\
&\leq \gamma.
\end{aligned}$$

Hence, we have the lemma. □

Lemma 5.9 For any $\gamma > 0$, there exists n_0 such that for any $n \geq n_0$,

$$E\left[\frac{1}{n} s_n(Z_1^n, \pi_n^L, L)\right] \leq S^*(\boldsymbol{X}) + \gamma \quad \forall \boldsymbol{Z} \in \Lambda(\boldsymbol{X}).$$

Proof: By (5.33), we can choose $k_{\boldsymbol{Z}}$ for each $\boldsymbol{Z} \in \Lambda(\boldsymbol{X})$ such that for any $k \geq k_{\boldsymbol{Z}}$,

$$\begin{aligned}
E\left[\frac{1}{k} s_k(Z_1^k, \pi_k^L, L)\right] &\leq \bar{S}(\boldsymbol{Z}) + \gamma/3 \\
&\leq S^*(\boldsymbol{X}) + \gamma/3.
\end{aligned}$$

By Lemma 5.8, there exists $\zeta = \zeta(k_{\boldsymbol{Z}}, \gamma/3)$ and an open neighborhood $U_{\boldsymbol{Z}} = \{\hat{\boldsymbol{Z}} : \nu(\boldsymbol{Z}, \hat{\boldsymbol{Z}}) < \zeta\}$ such that for any $\hat{\boldsymbol{Z}} \in U_{\boldsymbol{Z}}$,

$$E\left[\frac{1}{k_{\boldsymbol{Z}}} s_{k_{\boldsymbol{Z}}}(\hat{Z}_1^{k_{\boldsymbol{Z}}}, \pi_{k_{\boldsymbol{Z}}}^L, L)\right] \leq S^*(\boldsymbol{X}) + 2\gamma/3.$$

Moreover, Lemma 5.7 implies that there exists $n_{\boldsymbol{Z}} = n_{\boldsymbol{Z}}(\gamma/3, k_{\boldsymbol{Z}}, \mathcal{Y})$ such that for any $n \geq n_{\boldsymbol{Z}}$ and any $\hat{\boldsymbol{Z}} \in U_{\boldsymbol{Z}}$,

$$E\left[\frac{1}{n} s_n(\hat{\boldsymbol{Z}}_1^n, \pi_n^L, L)\right] \leq S^*(\boldsymbol{X}) + \gamma.$$

Since the open sets $U_{\boldsymbol{Z}}$ ($\boldsymbol{Z} \in \Lambda(\boldsymbol{X})$) cover $\Lambda(\boldsymbol{X})$ and $\Lambda(\boldsymbol{X})$ is compact (see [11] and references there in), we can choose finitely many $\boldsymbol{Z}(1), \cdots, \boldsymbol{Z}(t)$ such that the sets $U_{\boldsymbol{Z}(1)}, \cdots, U_{\boldsymbol{Z}(t)}$ also cover $\Lambda(\boldsymbol{X})$. Thus, $n_0 = \max_{1 \leq i \leq t} n_{\boldsymbol{Z}(i)}$ satisfies the lemma. □

Proof of Theorem 5.7: First, we prove (5.40). For any (π_k, σ_k),

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{k} E\left[s_k(X_{tk+1}^{(t+1)k}, \pi_k, \sigma_k)\right] \\
&\geq \limsup_{k\to\infty} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} E\left[\frac{1}{k} s_k(X_{tk+1}^{(t+1)k}, \pi_k^L, L)\right] && (5.42) \\
&= \limsup_{k\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\frac{1}{k} s_k(X_i^{i+k-1}, \pi_k^L, L)\right] && (5.43) \\
&\geq \inf_k \inf_{(\pi_k', \sigma_k')} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left[\frac{1}{k} s_k(X_i^{i+k-1}, \pi_k', \sigma_k')\right] \\
&= \inf_k \inf_{(\pi_k', \sigma_k')} \sup_{Z\in\Lambda(X)} E\left[\frac{1}{k} s_k(Z_1^k, \pi_k', \sigma_k')\right] && (5.44) \\
&\geq \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} \inf_k \inf_{(\pi_k', \sigma_k')} E\left[\frac{1}{k} s_k(Z_1^k, \pi_k', \sigma_k')\right] \\
&= \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} \liminf_{k\to\infty} \inf_{(\pi_k', \sigma_k')} E\left[\frac{1}{k} s_k(Z_1^k, \pi_k', \sigma_k')\right] && (5.45) \\
&= \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} \bar{S}(\boldsymbol{Z}) && (5.46) \\
&= S^*(\boldsymbol{X}),
\end{aligned}$$

where the equality (5.42) follows from Theorem 3.2, (5.43) from (5.15), (5.44) from the definition of the stationary hull $\Lambda(\boldsymbol{X})$, (5.45) from Lemma 5.6, and (5.46) from Theorem 5.6. Thus, we have (5.40).

Next, we prove (5.41). The encoders $\{(\pi_k^L, L)\}_{k=1}^{\infty}$ satisfy that

$$\begin{aligned}
&\lim_{k\to\infty} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{k} E\left[s_k(X_{tk+1}^{(t+1)k}, \pi_k^L, L)\right] \\
&= \inf_k \limsup_{n\to\infty} \inf_{(\pi_k', \sigma_k')} \frac{1}{n} \sum_{i=1}^{n} E\left[\frac{1}{k} s_k(X_i^{i+k-1}, \pi_k', \sigma_k')\right] && (5.47)
\end{aligned}$$

$$\begin{aligned}
&\le \inf_k \inf_{(\pi'_k,\sigma'_k)} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{1}{k}s_k(X_i^{i+k-1},\pi'_k,\sigma'_k)\right] \\
&= \inf_k \inf_{(\pi'_k,\sigma'_k)} \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} E\left[\frac{1}{k}s_k(Z_1^k,\pi'_k,\sigma'_k)\right] \\
&\le \limsup_{n\to\infty} \sup_{\boldsymbol{Z}\in\Lambda(\boldsymbol{X})} E\left[\frac{1}{n}s_n(Z_1^n,\pi_n^L,L)\right] \\
&\le S^*(\boldsymbol{X}), \qquad (5.48)
\end{aligned}$$

where the equality (5.47) follows from Theorem 5.2 and (5.48) from Lemma 5.9. Hence, we have (5.41). □

Chapter 6

Conclusions

In this thesis, we investigated the lossless and lossy coding problems of individual sequences. In the lossless coding problem, we clarified the relations among two kinds of empirical entropies, the self entropy rate, the Lempel-Ziv complexity, and the compressibility of individual sequences. In the fixed-rate, fixed-distortion, and fixed-slope lossy coding problems, we clarified that, for each problem, the optimal performance attainable via blockwise lossy encoders is characterized by using empirical distributions of sequences. The results are summarized as follows:

$$\begin{aligned} C(x) \leq \quad \rho^{bl}(x) &= \hat{H}(x) = \tilde{H}(x) = h(x), \\ \delta(R|x) &= \hat{D}(R|x) = \tilde{D}(R|x), \\ \rho(D|x) &= \hat{R}(D|x) = \tilde{R}(D|x), \\ \xi(x) &= \hat{S}(x) = \tilde{S}(x). \end{aligned}$$

The complexity $\hat{H}(x)$ (resp. $\hat{D}(R|x)$, $\hat{R}(D|x)$, $\hat{S}(x)$) gives the optimal performance $\rho^{bl}(x)$ (resp. $\delta(R|x)$, $\rho(D|x)$, $\xi(x)$) attainable by blockwise encoders. These results shows that when we consider coding problems of individual sequences, $\hat{H}(x)$ (resp. $\hat{D}(R|x)$, $\hat{R}(D|x)$, $\hat{S}(x)$) plays an important role analogous to the entropy rate $H(\boldsymbol{X})$ (resp. the distortion-rate function $D(R|\boldsymbol{X})$, the rate-distortion function $R(D|\boldsymbol{X})$, the optimal cost $S(\boldsymbol{X})$) of a stationary ergodic source. It should be pointed out that these complexities $\hat{H}(x)$, $\hat{D}(R|x)$, $\hat{R}(D|x)$, and $\hat{S}(x)$ are defined by using the empirical distribution $p_k(\cdot|x_1^n)$ of the sequence x. This fact indicates that each coding problem of an individual sequence x can be regarded as the coding problem of the source which is defined by the empirical distribution of the sequence. Hence, the ergodic theorem allows us to apply our results on the coding problems of individual sequences to the coding problems of stationary ergodic sources. In this thesis, it was shown that for almost all sequence x emitted from a stationary ergodic source $\boldsymbol{X}$, $\hat{H}(x)$ (resp. $\hat{D}(R|x)$, $\hat{R}(D|x)$, $\hat{S}(x)$) is equal to $H(\boldsymbol{X})$ ($D(R|\boldsymbol{X})$, $R(D|\boldsymbol{X})$, $S(\boldsymbol{X})$).

As an application of the above results, it was shown that a variation of LZ78 code is universal for a set A for which there exists countable length functions Σ such that $A \subset \bigcup_{\sigma \in \Sigma} \mathcal{X}^{\infty}(\sigma)$. Moreover, we proposed the fixed-rate, fixed-distortion, and fixed-slope universal lossy encoders based on the complexity of sequences. For each coding problem, the relation between the coding theorem of individual sequences and that of stationary ergodic sources was investigated.

There are some important problems to be solved in the further research:

1. While Corollary 2.1 shows a sufficient condition for the existence of a universal lossless encoder for a set of individual sequences, we should clarify a necessary condition. It should be also clarified that the assumptions introduced in lossy coding problems (for example, the existence of a reference letter) are necessary or not.

2. While we investigated the asymptotic performance of blockwise encoders, it is also important to evaluate the performance of block encoders when the block size k is fixed.

Appendix A

Source and Entropy

In this appendix, we summarize the quantities defined for random variables, and introduce some kinds of stochastic sources.

Let U be a random variable taking values in a finite or countably infinite alphabet $\mathcal{U}$ with the probabilistic distribution $P_U(u) = \Pr\{U = u\}$ $(u \in \mathcal{U})$. Then, the *entropy* of U is defined as

$$H(U) \stackrel{\triangle}{=} \sum_{u \in \mathcal{U}} P_U(u) \log \frac{1}{P_U(u)}.$$

Let (U, V) be a pair of random variables U and V such that U (resp. V) takes values in a finite or countably infinite alphabet $\mathcal{U}$ (resp. $\mathcal{V}$). Then, the *conditional entropy* is defined as

$$H(U|V) \stackrel{\triangle}{=} \sum_{(u,v) \in \mathcal{U} \times \mathcal{V}} P_{UV}(uv) \log \frac{1}{P_{U|V}(u|v)}.$$

Let $\mathcal{X}$ be arbitrary set. A sequence $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ of random variables on a set $\mathcal{X}$ is called a *source* with the alphabet $\mathcal{X}$. If $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ satisfies

$$P_{X_1^n}(x_1^n) = P_{X_i^{i+n-1}}(x_1^n) \quad \forall x_1^n, \forall i \in \mathcal{N},$$

then, $\boldsymbol{X}$ is called a *stationary source*. Moreover, a source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ is called an *ergodic source* if $\boldsymbol{X}$ satisfies that

$$G = \{\hat{x} : \hat{x}_i = x_{i+1}, x \in G\} \Rightarrow \Pr\{X_1^{\infty} \in G\} = 1 \text{ or } \Pr\{X_1^{\infty} \in G\} = 0,$$

for any $G \subset \mathcal{X}^{\infty}$. A source which is stationary and ergodic is called a *stationary ergodic source*. For a stationary ergodic source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ with finite or countably infinite alphabet $\mathcal{X}$, the *entropy rate* of $\boldsymbol{X}$ is given by

$$H(\boldsymbol{X}) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} H(X_1^n),$$

where it is known that the limit exists [2, 7].

A source $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$ is called a *k-th order Markov source* with the initial state $x_{1-k}^0 \in \mathcal{X}^k$ if

$$P_{X_1^n}(x_1^n) = \prod_{i=1}^{n-k+1} P_{X_{k+1}|X_1^k}(x_i|x_{i-k}^{i-1}),$$

holds for any $x_1^n \in \mathcal{X}^n$ and $n \in \mathcal{N}$. In other words, a k-th order Markov source is defined by the transition probability $P_{X_{k+1}|X_1^k}(\cdot|\cdot)$ and the initial state $x_{1-k}^0 \in \mathcal{X}^k$.

Bibliography

[1] G. J. Chaitin, "On the length of programs for computing binary sequences," *J. Assoc. Comput. Math.*, vol. 13, pp. 547–569, 1966.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[4] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 783–795, Nov. 1973.

[5] R. M. Gray, *Probability, Random Processes, and Ergodic Properties.* New York: Springer-Verlag, 1988.

[6] ——, *Entropy and Information Theory.* New York: Springer-Verlag, 1990.

[7] T. S. Han, *Information-spectrum methods in information theory.* New York: Springer-Verlag, 2002.

[8] T. S. Han and K. Kobayashi, *Mathematics of Information and Coding.* Tokyo, Japan: Baifukan, 1999, in Japanese.

[9] ——, *Mathematics of Information and Coding.* Boston, MA, USA: American Mathematical Society, 2001.

[10] J. C. Kieffer, "A unified approach to weak universal source coding," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 6, pp. 674–682, Nov. 1978.

[11] ——, "Fixed-rate encoding of nonstationary information sources," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 5, pp. 651–655, Sep. 1987.

[12] ——, "Sample converses in source coding theory," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 263–268, Mar. 1991.

[13] ——, "A survey of the theory of source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1473–1490, Sep. 1993.

[14] J. C. Kieffer and E.-h. Yang, "Grammar-based codes: a new class of universal lossless codes," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 737–754, May 2000.

[15] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *Probl. Inform. Thransm.*, vol. 1, pp. 4–7, 1965.

[16] S. Kuzuoka and T. Uyematsu, "Relationship among complexities of individual sequences over countable alphabet," *IEICE Trans. Fundamentals*, vol. E89-A, no. 7, pp. 2047–2055, Jul. 2006.

[17] ——, "Universal lossy coding for individual sequences based on complexity functions," in *Proc. of 2006 International Symposium on Information Theory and its Applications (ISITA2006)*, Seoul, Korea, Oct. 2006, pp. 290–295.

[18] ——, "Universal lossy coding for individual sequences based on complexity functions," *IEICE Trans. Fundamentals*, vol. E90-A, no. 2, pp. 491–503, Feb. 2007.

[19] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 75–81, Jan. 1976.

[20] J. Muramatsu and F. Kanaya, "Distortion-complexity and rate-distortion function," *IEICE Trans. Fundamentals*, vol. E77-A, no. 8, pp. 1224–1229, Aug. 1994.

[21] D. S. Ornstein and P. C. Shields, "Universal almost sure data compression," *Annals of Probability*, vol. 15, pp. 441–452, 1990.

[22] E. Plotnik, M. J. Weinberger, and J. Ziv, "Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the lempel-ziv algorithm," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 66–72, Jun. 1992.

[23] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July and October 1948.

[24] ——, “Coding theorems for a discrete source with a fidelity criterion,” *IRE National Convention Record, Part 4*, pp. 142–163, 1959.

[25] P. Shields, *The Ergodic Theory of Discrete Sample Paths.* American Mathematical Society, 1996.

[26] R. J. Solomonoff, “A formal theory of inductive inference,” *Inform. Contr.*, vol. 7, pp. 1–22, 224–254, 1964.

[27] T. Uyematsu and F. Kanaya, “Relations among three optimal criteria of universal coding of individual sequences,” in *Proc. of 2003 Shannon Theory Workshop (STW2003)*, Gunma, Japan, Sep. 2003, pp. 15–20.

[28] ——, “Asymptotical optimality of two variations of Lempel-Ziv codes for sources with countably infinite alphabet,” *IEICE Trans. Fundamentals*, vol. E89-A, no. 10, pp. 2459–2465, Oct. 2006.

[29] T. Uyematsu and N. Shiota, “Asymptotical optimality of grammar-based codes for nonstationary sources,” in *Proc. of 3rd Asian-European Workshop on Information Theory*, Kamogawa, Japan, Jun. 2003, pp. 49–52.

[30] K. Visweswariah, S. R. Kulkarni, and S. Verdú, “Universal coding of nonstationary sources,” *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1633–1637, Jul. 2000.

[31] E.-h. Yang and J. C. Kieffer, “Simple universal lossy data compression schemes derived from the Lempel-Ziv algorithm,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 239–245, Jan. 1996.

[32] E.-h. Yang and Z. Zhang, “Variable-rate trellis source encoding,” *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 586–608, Mar. 1999.

[33] E.-h. Yang, Z. Zhang, and T. Berger, “Fixed-slope universal lossy data compression,” *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1465–1476, Sep. 1997.

[34] J. Ziv, “Distortion-rate theory for individual sequences,” *IEEE Trans. Inform. Theory*, vol. IT-26, no. 2, pp. 137–143, Mar. 1980.

[35] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 3, pp. 337–343, May 1977.

[36] ——, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

[37] J. Ziv, "Coding of sources with unknown statistics–I: Probability of encoding error," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 384–389, May 1972.

[38] ——, "Coding of sources with unknown statistics–II: Distortion relative to a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 389–394, May 1972.

[39] ——, "Coding theorems for individual sequences," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 4, pp. 405–412, Jul. 1978.

Author's Publications Related to This Dissertation

Articles in Journals

1. S. Kuzuoka and T. Uyematsu "Relationship among complexities of individual sequences over countable alphabet," *IEICE Trans. Fundamentals,* vol. E89-A, No. 7, pp. 2047–2055, July 2006.

2. S. Kuzuoka and T. Uyematsu, "Universal lossy coding for individual sequences based on complexity functions," *IEICE Trans. Fundamentals,* vol. E90-A, No. 2, pp. 491–503, Feb. 2007.

Peer-Reviewed Articles in Conference Proceedings

1. S. Kuzuoka and T. Uyematsu, "Relationship among complexities of individual sequences over countable alphabet," in *Proc. of 2004 International Symposium on Information Theory and its Applications (ISITA2004),* pp. 1416–1421, Parma, Italy, Oct. 10–13, 2004.

2. S. Kuzuoka and T. Uyematsu, "Fixed-slope universal lossy coding for individual sequences," in *Proc. of 2006 IEEE Information Theory Workshop (ITW'06),* pp. 244–248, Chengdu, China, Oct. 22–26, 2006.

3. S. Kuzuoka and T. Uyematsu, "Universal lossy coding for individual sequences based on complexity functions," in *Proc. of 2006 International Symposium on Information Theory and its Applications (ISITA2006),* pp. 290–295, Seoul, Korea, Oct. 29–Nov. 1, 2006.

Miscellaneous

1. S. Kuzuoka and T. Uyematsu, "Universal lossy coding for individual sequences based on complexity functions," *IEICE Technical Report,* IT2005-86, pp. 129–134, March 2006.

2. S. Kuzuoka and T. Uyematsu, "Fixed-slope universal lossy coding for individual sequences,"*IEICE Technical Report,* IT2006-32, pp. 43–48, July 2006.

3. S. Kuzuoka and T. Uyematsu, "Fixed-slope universal lossy coding for nonstationary sources," in *Proc. of the 29th Symposium on Information Theory and Its Applications (SITA2006),* pp. 347–350, Hokkaido, Japan, Nov. 28–Dec. 1, 2006.