Encoding Irreducible Markovian Functions of Sources: An Application of Supremus Typicality

Sheng Huang, Mikael Skoglund, *Senior Member, IEEE*

Abstract

This paper establishes an achievability theorem for compressing irreducible Markov sources using linear coding over finite rings. Based on this theorem, it demonstrates that linear encoders over non-field rings can be equally optimal as their field counterparts for compressing irreducible Markov sources. Applying the linear coding technique, the problem of characterizing the achievable coding rate region of encoding some discrete Markovian function of several correlated sources is addressed. Coding rate regions achieved by linear encoders are presented. It is shown that linear encoders over non-field rings strictly outperform their field counterparts for encoding many functions. To be more precise, it is proved that the set of coding rates achieved by linear encoders over certain non-field ring is strictly larger than the one achieved by the field versions, regardless which finite field is considered.

From the point of view of proof techniques, the above achievability results are not direct extensions of corresponding results for i.i.d. sources [1] in which the proofs are built on the argument of traditional typical sequences from Shannon [2]. Instead, a new type of typicality for sequences, termed Supremus typical sequences, is introduced. The Asymptotically Equipartition Property and a generalized typicality lemma of Supremus typical sequences are proved. Compared to the traditional version, Supremus typicality allows us to derive more accessible results, while corresponding ones based on traditional typicality are often hard to analyze as demonstrated in the paper. This is one of the reasons why we introduce this new concept. The achievability results mentioned above are in fact applications of this Supremus typicality argument.

Index Terms

Discrete Function, Sources with Memory, Source Coding, Markov, Linear Coding, Finite Ring

I. Introduction

According to Csiszár [3], it is known that linear coding over finite fields achieves the optimal coding rates for all Slepian–Wolf data compression scenarios [4] (the result for binary field is also credited to Elias in [5]). The motivation for Csiszár's work includes "the fact that in some source network problems linear codes (over finite field) appear superior to others (cf. Körner and Marton [5])". Unfortunately, the problem whether linear coding over finite non-field rings can be equally optimal for Slepian–Wolf data compression remained open [6].

S. Huang and M. Skoglund are with the Communication Theory Lab, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, 10044, Sweden e-mail: (sheng.huang@ee.kth.se; skoglund@ee.kth.se).

This work was funded in part by the Swedish Research Council.

Due to the fact/weakness that a ring might possess a non-invertible element (with respect to its multiplicative operation), the analysis of [3] for the field case does not lead to an optimality conclusion when applied to the ring case. However, the authors' recent work [1], [7], [8] proves that, for any i.i.d. data compression scenario, there always exists linear encoders over finite non-field rings that achieve optimality, the Slepian–Wolf region [4]; moreover, it is demonstrated that the ring version can strictly dominate its field counterpart in many source network problems (more details on these problems are given later) which also motivates Csiszár's study (cf. [1]). Readers are kindly referred to [1] for more details.

This paper will focus on bringing those techniques to the context of Markovian, instead of i.i.d., sources. However, this generalization is not at all straightforward because a Markov process is no longer "symmetric" (in the sense that $\Pr\{X_1 = a, X_2 = b\} \neq \Pr\{X_1 = b, X_2 = a\}$ in general). The proofs involve the Markov coupling-uncoupling theorem and the concept of stochastic complement from Meyer [9]. We will look deeper into the stochastic behaviours characterized by the stochastic complements of reduced processes of Markov chains/sources. Based on that, the concept *Supremus typicality* is introduced. A Supremus typicality encoding-decoding trick is applied to obtain a related achievability theorem and demonstrate that linear coding over rings (in particular non-field rings) can also be optimal (cf. [10, Theorem 1] for the optimal achievable region). (Note: one can still use the classical typicality encoding-decoding technique to obtain achievability results. However, it is hard to draw a conclusion whether the achieved region is optimal in particular for Markov sources with transition matrixes that do not satisfy [11, (6)], since it associates with the entropy rate of another non-necessarily Markov process induced. To the best of our knowledge, there is no efficient method to evaluate the entropy rate in general. Thus, it is hard to assess the result to draw the same conclusion. Please see Section IV for more details.)

To demonstrate the advantages of linear coding over finite rings compared to its field counterpart. We then turn to a class of source network problems that partially motivate the studies on linear coding techniques (cf. [3], [1]). These problems are stated as the follows:

Problem 2 (Source Coding for Computing a Function of Sources with or without Memory[1])**.** Let S_t $(t \in \mathcal{S} = \{1, 2, \cdots, s\})$ be a *source* that randomly generates discrete data

$$\cdots, X_t^{(1)}, X_t^{(2)}, \cdots, X_t^{(n)}, \cdots,$$

where $X_t^{(n)}$ has a finite sample space $\mathscr{X}_t$ for all $n \in \mathbb{N}^+$. Given a discrete function $g : \mathscr{X} \to \mathscr{Y}$, where $\mathscr{X} = \prod_{t \in \mathcal{S}} \mathscr{X}_t$, what is the biggest region $\mathcal{R}[g] \subset \mathbb{R}^s$ satisfying, $\forall$ $(R_1, R_2, \cdots, R_s) \in \mathcal{R}[g]$ and $\forall$ $\epsilon > 0$, $\exists$ $N_0 \in \mathbb{N}^+$, such that, $\forall$ $n > N_0$, there exist s *encoders* $\phi_t : \mathscr{X}_t^n \to \left[1, 2^{nR_t}\right], t \in \mathcal{S}$, and one *decoder* $\psi : \prod_{t \in \mathcal{S}} \left[1, 2^{nR_t}\right] \to \mathscr{Y}^n$ with

$$\Pr\left\{\vec{g}\left(X_1^n, \cdots, X_s^n\right) \neq \psi\left[\phi_1\left(X_1^n\right), \cdots, \phi_s\left(X_s^n\right)\right]\right\} < \epsilon,$$

[1] The numerator, Problem 2, is such defined to avoid confusing with [1, Problem 1] referred latter.

where

$$X_t^n = \left[X_t^{(1)}, X_t^{(2)}, \cdots, X_t^{(n)}\right] \text{ and}$$
$$\vec{g}\left(X_1^n, \cdots, X_s^n\right) = \left[Y^{(1)}, Y^{(2)}, \cdots, Y^{(n)}\right]^t$$

with $Y^{(n)} = g\left(X_1^{(n)}, X_2^{(n)}, \cdots, X_s^{(n)}\right)$?

The region $\mathcal{R}[g]$ is called the *achievable coding rate region* for computing g. A rate tuple $\mathbf{R} \in \mathbb{R}^s$ is said to be *achievable* for computing g (or simply achievable) if $\mathbf{R} \in \mathcal{R}[g]$. A region $\mathcal{R} \subset \mathbb{R}^s$ is said to be *achievable* for computing g (or simply achievable) if $\mathcal{R} \subseteq \mathcal{R}[g]$.

Problem 2 is a generalization of [1, Problem 1] which considers only the special case that the process

$$\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots,$$

where $X^{(n)} = \left[X_1^{(n)}, X_2^{(n)}, \cdots, X_s^{(n)}\right]$, in Problem 2 is i.i.d.. Related work for this special scenario (i.i.d. scenario) includes: [4], [7] which considers the case that g is an identity function; [5], [12] where g is the binary sum; [13], [14] for conditions under which that $\mathcal{R}[g]$ is strictly larger than the Slepian–Wolf region; [15], [16], [17], [18], [19], [1], [8] for an arbitrary discrete function g. For other variations, readers are referred to [15], [20], [16], [17], [21], [22], [23], [24] for related problems including rate distortion, noisy channels and network coding scenarios.

We will investigate Problem 2 based on some additional Markovian constraints since, without any constraints on the stochastic behavior of the sources, the original scenario is too general to allow proper analysis. We henceforward assume that:

(c1) There exist some finite ring $\mathfrak{R}$, functions $k_t : \mathscr{X}_t \to \mathfrak{R}$ $(t \in \mathcal{S})$ and $h : \mathfrak{R} \to \mathscr{Y}$ with

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t \in \mathcal{S}} k_t(x_t)\right), \tag{1}$$

such that $\left\{\sum_{t \in \mathcal{S}} k_t\left(X_t^{(n)}\right)\right\}$ is irreducible[2] Markovian[3]. Thus, $\left\{g\left(X^{(n)}\right)\right\}$ is a Markovian function[4].

A linear and irreducible Markovian function is one of the many examples of such a function g satisfying (c1). By Lemma II.15 and Lemma B.1, (c1) includes:

(c0) g is arbitrary, while $\left\{X^{(n)}\right\}$ forms an irreducible Markov chain with transition matrix

$$\mathbf{P}_0 = c_1\mathbf{U} + (1 - c_1)\mathbf{1}, \tag{2}$$

where all rows of $\mathbf{U}$ are identical to some unitary vector $[u_x]_{x \in \mathscr{X}}$, $\mathbf{1}$ is an identity matrix and $0 \leq c_1 \leq 1$.

(c0) is very interesting because:

1) A stationary finite-state Markov chain $\left\{X^{(n)}\right\}$ admits a transition matrix of the form (2), if and only if $\left\{\Gamma\left(X^{(n)}\right)\right\}$ is Markovian for all feasible mappings Γ [11, Theorem 3];

[2] Irreducible is a more general condition compared to stationary ergodic which is (implicitly) assumed in some literature (cf. [25], [26]).

[3] For any finite discrete function g, such a finite ring $\mathfrak{R}$ and functions k_t's and h always exist by Lemma II.15. However, the Markovian condition is not guaranteed in general.

[4] A Markovian function is defined to be a Markov process that is a function of another arbitrary process [11].

2) $\left\{X^{(n)}\right\}$ is an i.i.d. process if $c_1 = 1$.

Making use of the linear coding technique introduced previously, we address Problem 2 of computing g regarding each of the previous conditions, (c0) and (c1). Inner bounds for $\mathcal{R}[g]$ are presented. It is demonstrated that the achievable regions given by these inner bounds are larger than Cover's region [10, Theorem 1]. Even more interestingly, our method (for computing some g) even works for some cases to which [10, Theorem 1] does not apply, because $\left\{X^{(n)}\right\}$ is not asymptotically mean stationary (a.m.s.)[5] [27]. Finally, a comparison between linear encoder over non-field ring and its field counterpart is carried out. It is seen that the non-field ring version offers many advantages, including strictly outperforming the field version in terms of achieving larger achievable region for computing (infinitely) many functions. In this sense, we conclude that linear coding over finite field is not optimal.

This paper is organized as the follows: Section II contains preliminaries used. Section III introduces the concept Supremus typical sequence and investigates related Asymptotically Equipartition Properties (AEP) and typicality lemmas. Section IV are dedicated to establish the achievability theorem of linear coding over finite rings for compressing Markov sources. Section V addresses Problem 2 as an application of the linear coding technique. Section VI is to demonstrate that linear coding over rings (non-field rings) is in strict upper hand compared to its field counterpart in many settings raised from Problem 2.

II. Preliminaries

Required concepts and properties are listed in this section to partially make the paper self-contained, at the same time, to clarify delicate aspects of concepts and (implicit) assumptions sometimes defined slightly differently in other literature. Readers are recommended to go thought (quickly) to identify our notation and universal assumptions.

A. *Index Oriented Matrix Operations*

Let $\mathscr{X}$, $\mathscr{Y}$ and $\mathscr{Z}$ be three countable sets with or without *orders* defined, e.g.

$$\mathscr{X} = \{(0,0),(0,1),(1,1),(1,0)\} \text{ and } \mathscr{Y} = \{\alpha,\beta\} \times \mathbb{N}^+.$$

In many places hereafter, we write $[p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}}$ ($[p_i]_{i\in\mathscr{X}}$) for a "matrix" ("vector") whose "(i,j)th" ("ith") entry is $p_{i,j}$ (p_i) $\in \mathbb{R}$. Matrices $\left[p'_{i,j}\right]_{i\in\mathscr{X},j\in\mathscr{Y}}$ and $[q_{j,k}]_{j\in\mathscr{Y},k\in\mathscr{Z}}$ are similarly defined. Let $\mathbf{P} = [p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}}$. For subsets $A \subseteq \mathscr{X}$ and $B \subseteq \mathscr{Y}$, $\mathbf{P}_{A,B}$ is designated for the "submatrix" $[p_{i,j}]_{i\in A,j\in B}$. We will use "index oriented" operations, namely

$$[p_i]_{i\in\mathscr{X}}[p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}} = \left[\sum_{i\in\mathscr{X}} p_i p_{i,j}\right]_{j\in\mathscr{Y}};$$

$$[p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}} + \left[p'_{i,j}\right]_{i\in\mathscr{X},j\in\mathscr{Y}} = \left[p_{i,j} + p'_{i,j}\right]_{i\in\mathscr{X},j\in\mathscr{Y}};$$

$$[p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}}[q_{j,k}]_{j\in\mathscr{Y},k\in\mathscr{Z}} = \left[\sum_{j\in\mathscr{Y}} p_{i,j} q_{j,k}\right]_{i\in\mathscr{X},k\in\mathscr{Z}}.$$

[5] To our best knowledge, asymptotically mean stationary is the most general condition known in literature based on which the Shannon–McMillon–Breiman Theorem holds [27].

In addition, a matrix $\mathbf{P}_{A,A} = [p_{i,j}]_{i,j \in A}$ is said to be an *identity matrix* if and only if $p_{i,j} = \delta_{i,j}$ (Kronecker delta), $\forall\, i, j \in A$. We often indicate an identity matrix with $\mathbf{1}$ whose size is known from the context, while designate $\mathbf{0}$ as the *zero matrix* (all of whose entries are 0) of size known from the context. For any matrix $\mathbf{P}_{A,A}$, its *inverse* (if exists) is some matrix $\mathbf{Q}_{A,A}$ such that $\mathbf{Q}_{A,A}\mathbf{P}_{A,A} = \mathbf{P}_{A,A}\mathbf{Q}_{A,A} = \mathbf{1}$. Let $[p_i]_{i \in \mathscr{X}}$ be non-negative and *unitary*, i.e. $\sum\limits_{i \in \mathscr{X}} p_i = 1$, and $[p_{i,j}]_{i \in \mathscr{X}, j \in \mathscr{Y}}$ be non-negative and $\sum\limits_{j \in \mathscr{Y}} p_{i,j} = 1$ (such a matrix is termed a *stochastic matrix*). For discrete random variables X and Y with sample spaces $\mathscr{X}$ and $\mathscr{Y}$, respectively, $X \sim [p_i]_{i \in \mathscr{X}}$ and $(X, Y) \sim [p_i]_{i \in \mathscr{X}} [p_{i,j}]_{i \in \mathscr{X}, j \in \mathscr{Y}}$ state for

$$\Pr\{X = i\} = p_i \text{ and } \Pr\{X = i, Y = j\} = p_i p_{i,j},$$

for all $i \in \mathscr{X}$ and $j \in \mathscr{Y}$, respectively.

B. Markov Chains and Strongly Markov Typical Sequences

If not specified, we assume that all Markov chains considered throughout this paper are *finite-state* and *homogeneous*. However, they are not necessarily *stationary* [28, pp. 71], or their *initial distributions* are unknown.

Definition II.1. Given a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}$ with state space $\mathscr{X}$, the *transition matrix* of $\mathscr{M}$ is defined to be the stochastic matrix $\mathbf{P} = [p_{i,j}]_{i,j \in \mathscr{X}}$, where $p_{i,j} = \Pr\left\{X^{(2)} = j \middle| X^{(1)} = i\right\}$. Moreover, $\mathscr{M}$ is said to be *irreducible* if and only if $\mathbf{P}$ is *irreducible*, namely, there exists no $\emptyset \neq A \subsetneq \mathscr{X}$ such that $\mathbf{P}_{A,A^c} = \mathbf{0}$.

Definition II.2. A state j of a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}$ is said to be *recurrent* if $\Pr\left\{T < \infty \middle| X^{(0)} = j\right\} = 1$, where $T = \inf\{n > 0 | X^{(n)} = j\}$. If in addition the conditional expectation $\mathbb{E}\{T | X^{(0)} = j\} < \infty$, then j is said to be *positive recurrent*. $\mathscr{M}$ is said to be *positive recurrent* if all states are positive recurrent.

Theorem II.3 (Theorem 1.7.7 of [29])**.** *An irreducible Markov chain $\mathscr{M}$ with state space $\mathscr{X}$ is positive recurrent, if and only if it admits a non-negative unitary vector $\pi = [p_j]_{j \in \mathscr{X}}$, such that $\pi\mathbf{P} = \pi$, where $\mathbf{P}$ is the transition matrix of $\mathscr{M}$. Moreover, π is unique and is called the* invariant (stationary) distribution.

Theorem II.4 (Theorem 2.31 of [30])**.** *A finite-state irreducible Markov chain is positive recurrent.*

Clearly, all irreducible Markov chains considered in this paper admit a unique invariant distribution (which is not necessarily the initial distribution), since they are assumed to be simultaneously finite-state and homogeneous (unless otherwise specified).

Definition II.5 (Strong Markov Typicality (cf. [25], [26]))**.** Let $\mathscr{M} = \left\{X^{(n)}\right\}$ be an irreducible Markov chain with state space $\mathscr{X}$, and $\mathbf{P} = [p_{i,j}]_{i,j \in \mathscr{X}}$ and $\pi = [p_j]_{j \in \mathscr{X}}$ be its transition matrix and invariant distribution, respectively. For any $\epsilon > 0$, a sequence $\mathbf{x} \in \mathscr{X}^n$ of *length* n (≥ 2) is said to be *strongly Markov ϵ-typical* with respect to $\mathbf{P}$ if

$$\left| \frac{N(i, j; \mathbf{x})}{N(i; \mathbf{x})} - p_{i,j} \right| < \epsilon \text{ and } \left| \frac{N(i; \mathbf{x})}{n} - p_i \right| < \epsilon, \forall\, i, j \in \mathscr{X},$$

where $N(i, j; \mathbf{x})$ is the occurrences of sub-sequence $[i, j]$ in $\mathbf{x}$ and $N(i; \mathbf{x}) = \sum\limits_{j \in \mathscr{X}} N(i, j; \mathbf{x})$. The set of all strongly Markov ϵ-typical sequences with respect to $\mathbf{P}$ in $\mathscr{X}^n$ is denoted by $\mathcal{T}_\epsilon(n, \mathbf{P})$ or $\mathcal{T}_\epsilon$ for simplicity.

Let $\mathbf{P}$ and π be some stochastic matrix and non-negative unitary vector. We define $H(\pi)$ and $H(\mathbf{P}|\pi)$ to be $H(X)$ and $H(Y|X)$, respectively, for jointly discrete random variables (X, Y) such that $X \sim \pi$ and $(X, Y) \sim \pi\mathbf{P}$.

Proposition II.6 (AEP of Strongly Markov Typicality[6])**.** *Let $\mathscr{M} = \left\{X^{(n)}\right\}$ be an irreducible Markov chain with state space $\mathscr{X}$, and $\mathbf{P} = [p_{i,j}]_{i,j\in\mathscr{X}}$ and $\pi = [p_j]_{j\in\mathscr{X}}$ be its transition matrix and invariant distribution, respectively. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\ \epsilon_0 > \epsilon > 0$, $\forall\ n > N_0$ and $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{T}_\epsilon(n, \mathbf{P})$,*

1) $\exp_2\left[-n\left(H(\mathbf{P}|\pi) + \eta\right)\right] < \Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} < \exp_2\left[-n\left(H(\mathbf{P}|\pi) - \eta\right)\right]$*;*
2) $\Pr\left\{\mathbf{X} \notin \mathcal{T}_\epsilon(n, \mathbf{P})\right\} < \eta$*, where* $\mathbf{X} = \left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right]$*; and*
3) $|\mathcal{T}_\epsilon(n, \mathbf{P})| < \exp_2\left[n\left(H(\mathbf{P}|\pi) + \eta\right)\right]$.

C. Rings, Ideals and Linear Mappings

Definition II.7. The tuple $[\mathfrak{R}, +, \cdot]$ is called a *ring* if the following criteria are met:

1) $[\mathfrak{R}, +]$ is an *Abelian group*;
2) There exists a *multiplicative identity* $1 \in \mathfrak{R}$, namely, $1 \cdot a = a \cdot 1 = a$, $\forall\ a \in \mathfrak{R}$;
3) $\forall\ a, b, c \in \mathfrak{R}$, $a \cdot b \in \mathfrak{R}$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;
4) $\forall\ a, b, c \in \mathfrak{R}$, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$.

We often write $\mathfrak{R}$ for $[\mathfrak{R}, +, \cdot]$ when the *operations* considered are known from the context. The operation "$\cdot$" is usually written by juxtaposition, ab for $a \cdot b$, for all $a, b \in \mathfrak{R}$.

A ring $[\mathfrak{R}, +, \cdot]$ is said to be *commutative* if $\forall\ a, b \in \mathfrak{R}$, $a \cdot b = b \cdot a$. In Definition II.7, the identity of the group $[\mathfrak{R}, +]$, denoted by 0, is called the *zero*. A ring $[\mathfrak{R}, +, \cdot]$ is said to be *finite* if the cardinality $|\mathfrak{R}|$ is finite, and $|\mathfrak{R}|$ is called the *order* of $\mathfrak{R}$. The set $\mathbb{Z}_q$ of integers modulo q is a commutative finite ring with respect to the *modular arithmetic*.

Definition II.8 (cf. [31])**.** The *characteristic* of a finite ring $\mathfrak{R}$ is defined to be the smallest positive integer m, such that $\sum_{j=1}^{m} 1 = 0$, where 0 and 1 are the zero and the multiplicative identity of $\mathfrak{R}$, respectively. The characteristic of $\mathfrak{R}$ is often denoted by Char($\mathfrak{R}$).

Remark 1. Clearly, Char$(\mathbb{Z}_q) = q$. For a finite field $\mathbb{F}$, Char$(\mathbb{F})$ is always the prime q_0 such that $|\mathbb{F}| = q_0^n$ for some integer n [32, Proposition 2.137].

Definition II.9. A subset $\mathfrak{I}$ of a ring $[\mathfrak{R}, +, \cdot]$ is said to be a *left ideal* of $\mathfrak{R}$, denoted by $\mathfrak{I} \leq_l \mathfrak{R}$, if and only if

1) $[\mathfrak{I}, +]$ is a subgroup of $[\mathfrak{R}, +]$;
2) $\forall\ x \in \mathfrak{I}$ and $\forall\ r \in \mathfrak{R}$, $r \cdot x \in \mathfrak{I}$.

If condition 2) is replaced by

3) $\forall\ x \in \mathfrak{I}$ and $\forall\ r \in \mathfrak{R}$, $x \cdot r \in \mathfrak{I}$,

[6] Similar statements in many literature (cf. [25], [26]) assume that the Markov chain is stationary ergodic. The result is easy to generalize to irreducible Markov chain. To be rigorous, we include a proof in Appendix A.

then $\mathfrak{I}$ is called a *right ideal* of $\mathfrak{R}$, denoted by $\mathfrak{I} \leq_r \mathfrak{R}$. $\{0\}$ is a *trivial* left (right) ideal, usually denoted by 0.

It is well-known that if $\mathfrak{I} \leq_l \mathfrak{R}$, then $\mathfrak{R}$ is divided into disjoint *cosets* of equal size (cardinality). $|\mathfrak{I}|$ is called the *order* of $\mathfrak{I}$ if it is finite. For any coset $\mathfrak{J}$, $\mathfrak{J} = x + \mathfrak{I} = \{x + y | y \in \mathfrak{I}\}$, $\forall\ x \in \mathfrak{J}$. The set of all cosets, denoted by $\mathfrak{R}/\mathfrak{I}$, forms a *left module*. Similarly, $\mathfrak{R}/\mathfrak{I}$ becomes a *right module* if $\mathfrak{I} \leq_r \mathfrak{R}$. $\mathfrak{R}/\mathfrak{I}$ is also seen as a partition of $\mathfrak{R}$ [32, Ch. 1.6 and Ch. 2.9].

Example II.10. *Let* $M_{L,2} = \left\{ \begin{bmatrix} a & 0 \\ b & a \end{bmatrix} \middle| a, b \in \mathbb{Z}_2 \right\}$. $M_{L,2}$ *is a ring with respect to usual matrix addition and multiplication (note:* $M_{L,2}$ *is not isomorphic to* $\mathbb{Z}_2 \times \mathbb{Z}_2$*). Its only left ideal is* $\mathfrak{I} = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$. $M_{L,2}/\mathfrak{I} = \{\mathfrak{I}, \mathfrak{J}\}$, *where* $\mathfrak{J} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mathfrak{I} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \mathfrak{I}$.

Definition II.11. A mapping $f : \mathfrak{R}^n \to \mathfrak{R}^m$ given as:

$$\begin{gathered} f(x_1, x_2, \cdots, x_n) = \left(\sum_{j=1}^n a_{1,j} x_j, \cdots, \sum_{j=1}^n a_{m,j} x_j \right)^t \\ \left(f(x_1, x_2, \cdots, x_n) = \left(\sum_{j=1}^n x_j a_{1,j}, \cdots, \sum_{j=1}^n x_j a_{m,j} \right)^t \right), \\ \forall\ (x_1, x_2, \cdots, x_n) \in \mathfrak{R}^n, \end{gathered}$$

where $a_{i,j} \in \mathfrak{R}$ for all feasible i and j, is called a *left* (*right*) *linear mapping* over ring $\mathfrak{R}$. If $m = 1$, then f is called a *left* (*right*) *linear function* over $\mathfrak{R}$. The matrix $\mathbf{A} = [a_{i,j}]_{1 \leq i,j \leq n}$ is called the *coefficient matrix* of f.

In our later discussions, we mainly use left linear mappings (functions, encoders). They are simply referred to as linear mappings (functions, encoders). This will not give rise to confusion because left linearity and right linearity can always be distinguished from the context.

Lemma II.12 ([1])**.** *Let* $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^n$ *be two distinct sequences, where* $\mathfrak{R}$ *is a finite ring, and assume that* $\mathbf{y} - \mathbf{x} = (a_1, a_2, \cdots, a_n)^t$. *If* $f : \mathfrak{R}^n \to \mathfrak{R}^k$ *is a random linear mapping chosen uniformly at random, i.e. generate the* $k \times n$ *coefficient matrix* $\mathbf{A}$ *of* f *by independently choosing each entry of* $\mathbf{A}$ *from* $\mathfrak{R}$ *uniformly at random, then*

$$\Pr\{f(\mathbf{x}) = f(\mathbf{y})\} = |\mathfrak{I}|^{-k}, \tag{3}$$

where $\mathfrak{I}$ *denotes the left ideal* $\left\{ \sum_{i=1}^n r_i a_i \middle| r_i \in \mathfrak{R}, \forall\ 1 \leq i \leq n \right\}$.

D. Polynomial Functions

Definition II.13. A *polynomial function* of k *variables* over a finite ring $\mathfrak{R}$ is a function $g : \mathfrak{R}^k \to \mathfrak{R}$ of the form

$$g(x_1, x_2, \cdots, x_k) = \sum_{j=0}^m a_j x_1^{m_{1j}} x_2^{m_{2j}} \cdots x_k^{m_{kj}}, \tag{4}$$

where $a_j \in \mathfrak{R}$ and m and m_{ij}'s are non-negative integers. The set of all the polynomial functions of k variables over ring $\mathfrak{R}$ is designated by $\mathfrak{R}[k]$.

Remark 2. *Polynomial* and polynomial function are sometimes only defined over a commutative ring [32]. It is a very delicate matter to define them over a non-commutative ring [33], [34], due to the fact that x_1x_2 and x_2x_1 can become different objects. We choose to define "polynomial functions" with formula (4) because those functions are within the scope of this paper's interest.

Lemma II.14 ([1]). *For any discrete function $g : \prod_{i=1}^{k} \mathscr{X}_i \to \mathscr{Y}$ with $\mathscr{X}_i$'s and $\mathscr{Y}$ being finite, there always exist a finite ring (field) and a polynomial function $\hat{g} \in \mathfrak{R}[k]$ such that*

$$\nu \left(g \left(x_1, x_2, \cdots, x_k \right) \right) = \hat{g} \left(\mu_1(x_1), \mu_2(x_2), \cdots, \mu_k(x_k) \right)$$

for some injections $\mu_i : \mathscr{X}_i \to \mathfrak{R}$ ($1 \leq i \leq k$) and $\nu : \mathscr{Y} \to \mathfrak{R}$.

The important message conveyed by Lemma II.14 says that any discrete function defined on a finite domain is essentially a *restriction* [14, Definition II.3] of some polynomial function. Therefore, we can restrict the consideration of Problem 2 to all polynomial functions. This polynomial approach [14], [18] offers a very good insight into the general problem. After all, the algebraic structure of a polynomial function is much more accessible than that of an arbitrary mapping (function). Most importantly, a polynomial function can often be expressed in several formats. Some of them are very helpful in tackling Problem 2 [14], [18].

Lemma II.15 ([1]). *Let $\mathscr{X}_1, \mathscr{X}_2, \cdots, \mathscr{X}_s$ and $\mathscr{Y}$ be some finite sets. For any discrete function $g : \prod_{t=1}^{s} \mathscr{X}_t \to \mathscr{Y}$, there exist a finite ring (field) $\mathfrak{R}$, functions $k_t : \mathscr{X}_t \to \mathfrak{R}$ and $h : \mathfrak{R} \to \mathscr{Y}$, such that*

$$g(x_1, x_2, \cdots, x_s) = h \left(\sum_{t=1}^{s} k_t(x_t) \right). \tag{5}$$

We often name the polynomial function $\hat{g}$ in Lemma II.14 a *polynomial presentation* of g. The left hand side of (5) is termed a *nomographic function* (by terminology borrowed from [35]) over $\mathfrak{R}$. It is said to be a *nomographic presentation* of g. Readers are kindly referred to [18] for other interested formats. As a simple demonstration [14], one can see that the function $\min\{x, y\}$ defined on $\{0, 1\} \times \{0, 1\}$ (with order $0 < 1$) admits polynomial presentations $xy \in \mathbb{Z}_2[2]$ and $x + y - (x + y)^2$ defined on $\{0, 1\} \times \{0, 1\} \subsetneq \mathbb{Z}_3^2$. The second one gives a nomographic presentation.

III. Supremus Typical Sequences

This paper will from now on not rely on the traditional (weakly/strongly) typical sequence argument [2]. Instead, a new typicality concept is defined. This new concept is stronger in the sense of characterizing the stochastic behaviors of random processes/sources. Although this concept is only defined for and applied to Markovian processes/sources in this paper, the idea can be easily generalized to other random processes/sources, e.g. a.m.s. processes/sources [27]. Before proceeding, we need the following background material from Meyer [9], cited here for completeness.

Given a Markov chain $\mathscr{M} = \left\{ X^{(n)} \right\}$ with state space $\mathscr{X}$ and a non-empty subset A of $\mathscr{X}$, let

$$T_{A,l} = \begin{cases} \inf \left\{ n > 0 | X^{(n)} \in A \right\}; & l = 1, \\ \inf \left\{ n > T_{A,l-1} | X^{(n)} \in A \right\}; & l > 1, \\ \sup \left\{ n < T_{A,l+1} | X^{(n)} \in A \right\}; & l < 1. \end{cases}$$

It is well-known that $\mathscr{M}_A = \left\{X^{(T_{A,l})}\right\}$ is Markov by the strong Markov property [29, Theorem 1.4.2]. In particular, if $\mathscr{M}$ is irreducible, so is $\mathscr{M}_A$. To be more precise, if $\mathscr{M}$ is irreducible, and write its invariant distribution and transition matrix as $\pi = [p_i]_{i \in \mathscr{X}}$ and

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{A,A} & \mathbf{P}_{A,A^c} \\ \mathbf{P}_{A^c,A} & \mathbf{P}_{A^c,A^c} \end{bmatrix},$$

respectively, then

$$\mathbf{S}_A = \mathbf{P}_{A,A} + \mathbf{P}_{A,A^c} \left(\mathbf{1} - \mathbf{P}_{A^c,A^c}\right)^{-1} \mathbf{P}_{A^c,A},$$

is the transition matrix of $\mathscr{M}_A$ [9, Theorem 2.1 and Section 3]. $\pi_A = \left[\frac{p_i}{\sum_{j \in A} p_j}\right]_{i \in A}$ is an invariant distribution of $\mathbf{S}_A$, i.e. $\pi_A \mathbf{S}_A = \pi_A$ [9, Theorem 2.2]. Since $\mathscr{M}_A$ inherits irreducibility from $\mathscr{M}$ [9, Theorem 2.3], π_A is unique. The matrix $\mathbf{S}_A$ is termed the *stochastic complement* of $\mathbf{P}_{A,A}$ in $\mathbf{P}$, while $\mathscr{M}_A$ is named a *reduced Markov chain* (or *reduced process*) of $\mathscr{M}$. It has state space A obviously.

Definition III.1 (Supremus Typicality)**.** Following the notation defined above, given $\epsilon > 0$ and a sequence $\mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathscr{X}^n$ of length n ($\geq 2\,|\mathscr{X}|$), let $\mathbf{x}_A$ be the subsequence of $\mathbf{x}$ formed by all those $x^{(l)}$'s that belong to A in the original ordering. $\mathbf{x}$ is said to be *Supremus ϵ-typical* with respect to $\mathbf{P}$, if and only if $\mathbf{x}_A$ is strongly Markov ϵ-typical with respect to $\mathbf{S}_A$ for any feasible non-empty subset A of $\mathscr{X}$.

In Definition III.1, the set of all Supremus ϵ-typical sequences with respect to $\mathbf{P}$ in $\mathscr{X}^n$ is denoted as $\mathcal{S}_\epsilon(n, \mathbf{P})$ or $\mathcal{S}_\epsilon$ for simplicity. $\mathbf{x}_A$ is called a *reduced subsequence* (with respect to A) of $\mathbf{x}$. It follows immediately form the definition that

Proposition III.2. *Every reduced subsequence of a Supremus ϵ-typical sequence is Supremus ϵ-typical.*

However, the above proposition does not hold for strongly Markov ϵ-typical sequences. Namely, *a reduced subsequence of a strongly Markov ϵ-typical sequence is not necessarily strongly Markov ϵ-typical.*

Example III.3. Let $\{\alpha, \beta, \gamma\}$ be the state space of an i.i.d. process with a uniform distribution, i.e.

$$\mathbf{P} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix},$$

and

$$\mathbf{x} = (\alpha, \beta, \gamma, \alpha, \beta, \gamma, \alpha, \beta, \gamma).$$

It is easy to verify that $\mathbf{x}$ is a strongly Markov $5/12$-typical sequence. However, the reduced subsequence

$$\mathbf{x}_{\{\alpha,\gamma\}} = (\alpha, \gamma, \alpha, \gamma, \alpha, \gamma)$$

is no long a strongly Markov $5/12$-typical sequence, because $\mathbf{S}_{\{\alpha,\gamma\}} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ and

$$\left|\frac{\text{the number of subsequence } (\alpha, \alpha)\text{'s in } \mathbf{x}_{\{\alpha,\gamma\}}}{6} - 0.5\right| = |0 - 0.5| > \frac{5}{12}.$$

Proposition III.4 (AEP of Supremus Typicality)**.** *Let $\mathscr{M} = \left\{X^{(n)}\right\}$ be an irreducible Markov chain with state space $\mathscr{X}$, and $\mathbf{P} = [p_{i,j}]_{i,j\in\mathscr{X}}$ and $\pi = [p_j]_{j\in\mathscr{X}}$ be its transition matrix and invariant distribution, respectively. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\ \epsilon_0 > \epsilon > 0$, $\forall\ n > N_0$ and $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P})$,*

1) $\exp_2\left[-n\left(H(\mathbf{P}|\pi) + \eta\right)\right] < \Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} < \exp_2\left[-n\left(H(\mathbf{P}|\pi) - \eta\right)\right]$*;*
2) $\Pr\left\{\mathbf{X} \notin \mathcal{S}_\epsilon(n, \mathbf{P})\right\} < \eta$*, where* $\mathbf{X} = \left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right]$*; and*
3) $|\mathcal{S}_\epsilon(n, \mathbf{P})| < \exp_2\left[n\left(H(\mathbf{P}|\pi) + \eta\right)\right]$.

Proof: Note that $\mathcal{T}_\epsilon(n, \mathbf{P}) \supseteq \mathcal{S}_\epsilon(n, \mathbf{P})$. Thus, 1) and 3) are inherited from the AEP of strongly Markov typicality. In addition, 2) can be proved without any difficulty since any reduced Markov chain of $\mathscr{M}$ is irreducible and the number of reduced Markov chains of $\mathscr{M}$ is, $2^{|\mathscr{X}|} - 1$, finite. ■

Remark 3. Motivated by Definition III.1, Proposition III.4 and two related typicality lemmas in Appendix C, one can define the concept of *Supremus type* resembling other classic notions of types [26], e.g. Markov type [25].

Remark 4. It is known that Shannon's (weakly/strongly) typical sequences [2] are defined to be those sequences "representing" the stochastic behavior of the whole random process. To be more precise, a non (weakly/strongly) typical sequence is unlikely to be produced by the random procedure (Proposition II.6). However, the study of induced transformations[7] in ergodic theory [37] suggests that (weakly/strongly) typical sequences that are not Supremus typical form also a low probability set [36]. When the random procedure propagates, it is highly likely that all reduced subsequences of the generated sequence also admit empirical distributions "close enough" to the genuine distributions of corresponding reduced processes as seen from Proposition III.4. Therefore, Supremus typical sequences "represent" the random process better. This difference has been seen from Proposition III.2 and Example III.3 and will be seen again in comparing the two typicality lemmas, Lemma III.5 and Lemma III.6, given later.

The following two typicality lemmas of Supremus typical sequences are the ring versions tailored for our discussions from the two given in Appendix C, respectively. From these two lemmas, we will start to see the impact brought to the analytic results by the differences between classical typicality and Supremus typicality.

Lemma III.5. *Let $\mathfrak{R}$ be a finite ring, $\mathscr{M} = \left\{X^{(n)}\right\}$ be an irreducible Markov chain whose state space, transition matrix and invariant distribution are $\mathfrak{R}$, $\mathbf{P}$ and $\pi = [p_j]_{j\in\mathfrak{R}}$, respectively. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\ \epsilon_0 > \epsilon > 0$, $\forall\ n > N_0$, $\forall\ \mathbf{x} \in \mathcal{S}_\epsilon(n, \mathbf{P})$ and $\forall\ \mathfrak{I} \leq_l \mathfrak{R}$,*

$$|S_\epsilon(\mathbf{x}, \mathfrak{I})| < \exp_2\left\{n\left[\sum_{A\in\mathfrak{R}/\mathfrak{I}}\sum_{j\in A} p_j H(\mathbf{S}_A|\pi_A) + \eta\right]\right\} \tag{6}$$

$$= \exp_2\left\{n\left[H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi) + \eta\right]\right\} \tag{7}$$

[7] See [36] for the correspondence between an induced transformation and a reduced process of a random process (a dynamical system).

where

$$S_\epsilon(\mathbf{x},\mathfrak{I}) = \{\mathbf{y}\in\mathcal{S}_\epsilon(n,\mathbf{P})|\,\mathbf{y}-\mathbf{x}\in\mathfrak{I}^n\},$$

$\mathbf{S}_A$ *is the stochastic complement of* $\mathbf{P}_{A,A}$ *in* $\mathbf{P}$*,* $\pi_A = \left[\frac{p_i}{\sum_{j\in A} p_j}\right]_{i\in A}$ *is the invariant distribution of* $\mathbf{S}_A$ *and*

$$\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} = \mathrm{diag}\left\{\{\mathbf{S}_A\}_{A\in\mathfrak{R}/\mathfrak{I}}\right\}.$$

Remark 5. By definition, for any $\mathbf{y}\in S_\epsilon(\mathbf{x},\mathfrak{I})$ in Lemma III.5, we have that $\mathbf{y}$ and $\mathbf{x}$ follow the same sequential pattern, i.e. the ith coordinates of both sequences are from the same coset of $\mathfrak{I}$. If $\mathfrak{I}=\mathfrak{R}$, then $S_\epsilon(\mathbf{x},\mathfrak{I})$ is the whole set of Supremus typical sequences. It is well-known that evaluating the cardinality of the set of all the (weakly/strongly) typical sequences is of great importance to the achievability part of the source coding theorem [4]. We will see from the next section that determining the number of (weakly/strongly/Supremus) typical sequences of certain sequential pattern is also very important to the achievability result for linear coding over finite rings.

Proof of Lemma III.5: Assume that $\mathbf{x}=\left[x^{(1)},x^{(2)},\cdots,x^{(n)}\right]$ and let $\mathbf{x}_A$ be the subsequence of $\mathbf{x}$ formed by all those $x^{(l)}$'s that belong to $A\in\mathfrak{R}/\mathfrak{I}$ in the original ordering. For any $\mathbf{y}=\left[y^{(1)},y^{(2)},\cdots,y^{(n)}\right]\in S_\epsilon(\mathbf{x},\mathfrak{I})$, obviously $y^{(l)}\in A$ if and only if $x^{(l)}\in A$ for all $A\in\mathfrak{R}/\mathfrak{I}$ and $1\le l\le n$. Let $\mathbf{x}_A=\left[x^{(n_1)},x^{(n_2)},x^{(n_{m_A})}\right]$ (note: $\sum_{A\in\mathfrak{R}/\mathfrak{I}} m_A = n$ and $\left|\frac{m_A}{n}-\sum_{j\in A}p_j\right| < |A|\,\epsilon+\frac{1}{n}$). By Proposition III.2, $\mathbf{y}_A=\left[y^{(n_1)},y^{(n_2)},y^{(n_{m_A})}\right]\in A^{m_A}$ is a strongly Markov ϵ-typical sequence of length m_A with respect to $\mathbf{S}_A$, since $\mathbf{y}$ is Supremus ϵ-typical. Additionally, by Proposition II.6, there exist $\epsilon_A>0$ and positive integer M_A such that the number of strongly Markov ϵ-typical sequences of length m_A is upper bounded by $\exp_2\{m_A\left[H(\mathbf{S}_A|\pi_A)+\eta/2\right]\}$ if $0<\epsilon<\epsilon_A$ and $m_A>M_A$. Therefore, if $0<\epsilon<\min_{A\in\mathfrak{R}/\mathfrak{I}}\epsilon_A$, $n>M=\max_{A\in\mathfrak{R}/\mathfrak{I}}\left\{\frac{1+M_A}{\left|\sum_{j\in A}p_j-|A|\,\epsilon\right|}\right\}$ (this guarantees that $m_A>M_A$ for all $A\in\mathfrak{R}/\mathfrak{I}$), then

$$\begin{aligned}|S_\epsilon(\mathbf{x},\mathfrak{I})| \le& \exp_2\left\{\sum_{A\in\mathfrak{R}/\mathfrak{I}} m_A\left[H(\mathbf{S}_A|\pi_A)+\eta/2\right]\right\}\\ =&\exp_2\left\{n\left[\sum_{A\in\mathfrak{R}/\mathfrak{I}}\frac{m_A}{n}H(\mathbf{S}_A|\pi_A)+\eta/2\right]\right\}.\end{aligned}$$

Furthermore, choose $0<\epsilon_0\le\min_{A\in\mathfrak{R}/\mathfrak{I}}\epsilon_A$ and $N_0\ge M$ such that $\frac{m_A}{n}<\sum_{j\in A}p_j+\frac{\eta}{2\sum_{A\in\mathfrak{R}/\mathfrak{I}}H(\mathbf{S}_A|\pi_A)}$ for all $0<\epsilon<\epsilon_0$ and $n>N_0$ and $A\in\mathfrak{R}/\mathfrak{I}$, we have

$$|S_\epsilon(\mathbf{x},\mathfrak{I})| <\exp_2\left\{n\left[\sum_{A\in\mathfrak{R}/\mathfrak{I}}\sum_{j\in A}p_jH(\mathbf{S}_A|\pi_A)+\eta\right]\right\},$$

(6) is established. Direct calculation yields (7). ■

At this point, one might argue to replace $S_\epsilon(\mathbf{x},\mathfrak{I})$ in Lemma III.5 with $T_\epsilon(\mathbf{x},\mathfrak{I})=\{\mathbf{y}\in\mathcal{T}_\epsilon(n,\mathbf{P})|\,\mathbf{y}-\mathbf{x}\in\mathfrak{I}^n\}$, the set of strongly Markov ϵ-typical sequences having the same sequential pattern as those from $S_\epsilon(\mathbf{x},\mathfrak{I})$, to keep the argument inside the classical typicality framework. Unfortunately, a reduced subsequence of a sequence from $T_\epsilon(\mathbf{x},\mathfrak{I})$ is not necessarily strongly Markov ϵ-typical anymore (Proposition III.2 fails). Thus, the same proof does not follow. Even though a corresponding bound of $|T_\epsilon(\mathbf{x},\mathfrak{I})|$ (see Lemma III.6) can be obtained, this bound is often very hard to evaluate as seen later.

Lemma III.6. *In Lemma III.5,*

$$|S_\epsilon(\mathbf{x},\mathfrak{I})| \leq |T_\epsilon(\mathbf{x},\mathfrak{I})| < \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty}\frac{1}{m}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) + \eta\right]\right\}, \tag{8}$$

where $Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} = X^{(m)} + \mathfrak{I}$ *is a random variable with sample space* $\mathfrak{R}/\mathfrak{I}$*.*

Proof: $|S_\epsilon(\mathbf{x},\mathfrak{I})| \leq |T_\epsilon(\mathbf{x},\mathfrak{I})|$ is obvious. Assume that $\mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right]$ and let

$$\overline{\mathbf{y}} = \left[x^{(1)} + \mathfrak{I}, x^{(2)} + \mathfrak{I}, \cdots, x^{(n)} + \mathfrak{I}\right].$$

For any $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in T_\epsilon(\mathbf{x},\mathfrak{I})$, obviously $y^{(l)} \in A$ if and only if $x^{(l)} \in A$ for all $A \in \mathfrak{R}/\mathfrak{I}$ and $1 \leq l \leq n$. Moreover,

$$\overline{\mathbf{y}} = \left[y^{(1)} + \mathfrak{I}, y^{(2)} + \mathfrak{I}, \cdots, y^{(n)} + \mathfrak{I}\right].$$

$\mathbf{y}$ is *jointly typical* with $\overline{\mathbf{y}}$ [10] with respect to the process

$$\cdots, \begin{pmatrix} X^{(1)} \\ Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \end{pmatrix}, \begin{pmatrix} X^{(2)} \\ Y_{\mathfrak{R}/\mathfrak{I}}^{(2)} \end{pmatrix}, \cdots, \begin{pmatrix} X^{(n)} \\ Y_{\mathfrak{R}/\mathfrak{I}}^{(n)} \end{pmatrix}, \cdots$$

Therefore, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\ \epsilon_0 > \epsilon > 0$ and $\forall\ n > N_0$,

$$\begin{aligned}|T_\epsilon(\mathbf{x},\mathfrak{I})| <& \exp_2\left\{n\left[\lim_{m\to\infty}\frac{1}{m}H\left(X^{(m)}, X^{(m-1)}, \cdots, X^{(1)}\right)\right.\right.\\ &\left.\left. - \lim_{m\to\infty}\frac{1}{m}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) + \eta\right]\right\}\\ =& \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty}\frac{1}{m}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) + \eta\right]\right\},\end{aligned}$$

where the equality follows from the fact that $\lim_{m\to\infty}\frac{1}{m}H\left(X^{(m)}, X^{(m-1)}, \cdots, X^{(1)}\right) = H\left(\mathbf{P}\left|\pi\right.\right)$ since $\mathscr{M}$ is irreducible Markov. ■

Remark 6. If $\mathfrak{R}$ in Lemma III.5 is a field, then both (7) and (8) are equivalent to

$$|S_\epsilon(\mathbf{x},\mathfrak{I})| < \exp_2\left[n\left(H\left(\mathbf{P}|\pi\right) + \eta\right)\right].$$

Or, if $\mathscr{M}$ in Lemma III.5 is i.i.d., then both (7) and (8) are equivalent to

$$|S_\epsilon(\mathbf{x},\mathfrak{I})| < \exp_2\left[n\left(H\left(X^{(1)}\right) - H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) + \eta\right)\right],$$

which is a special case of the *generalized conditional typicality lemma* [1, Lemma III.5].

Remark 7. In Lemma III.6, if $\mathbf{P} = c_1\mathbf{U} + (1-c_1)\mathbf{1}$ with all rows of $\mathbf{U}$ being identical and $0 \leq c_1 \leq 1$, then $\mathscr{M}' = \left\{Y_{\mathfrak{R}/\mathfrak{I}}^{(n)}\right\}$ is Markovian by Lemma B.1. As a conclusion,

$$\begin{aligned}&|S_\epsilon(\mathbf{x},\mathfrak{I})| \leq |T_\epsilon(\mathbf{x},\mathfrak{I})|\\ <& \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}\middle|Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}\right) + \eta\right]\right\}\\ =& \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - H\left(\mathbf{P}'|\pi'\right) + \eta\right]\right\},\end{aligned}$$

where $\mathbf{P}'$ and π' are the transition matrix and the invariant distribution of $\mathscr{M}'$ that can be easily calculated from $\mathbf{P}$.

From Remark 6 and Remark 7, we have seen that the two bounds (7) and (8) coincide, and both can be easily calculated for some special scenarios. Unfortunately, for general settings (when the initial distribution of $\mathscr{M}$ is not known or $\mathbf{P} \neq c_1\mathbf{U} + (1 - c_1)\mathbf{1}$ for any $\mathbf{U}$ and c_1), (8) becomes almost unaccessible because there is no efficient way to evaluate the entropy rate of $\left\{ Y^{(n)}_{\mathfrak{R}/\mathfrak{I}} \right\}$. On the other hand, (7) is always as straightforward as calculating the conditional entropy.

Example III.7. Let $\mathscr{M}$ be an irreducible Markov chain with state space $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Its transition matrix $\mathbf{P} = [p_{i,j}]_{i,j \in \mathbb{Z}_4}$ is given as the follows.

	0	1	2	3
0	.2597	.2093	.2713	.2597
1	.1208	.0872	.6711	.1208
2	.0184	.2627	.4101	.3088
3	.0985	.1823	.2315	.4877

(9)

Let $\mathfrak{I} = \{0, 2\}$. Notice that the initial distribution is unknown, neither $\mathbf{P} = c_1\mathbf{U} + (1 - c_1)\mathbf{1}$ for any $\mathbf{U}$ and c_1. Thus, the upper bound of $|S_\epsilon(\mathbf{x}, \mathfrak{I})|$ and $|T_\epsilon(\mathbf{x}, \mathfrak{I})|$ from (8) is not very meaningful for calculation since the entropy rate is not explicitly known. In contrast, we have that

$$|S_\epsilon(\mathbf{x}, \mathfrak{I})| < 2^{0.8791}$$

by (7).

The above is partially the reason we forsake the traditional (weakly/strongly) typical sequence argument of Shannon [2], and introduce an argument based on Supremus typicality.

IV. Achievability Theorem of Markov Source Compression with Linear Coding

Equipped with the foundation laid down by Proposition III.4, Lemma III.5 and Lemma III.6, we resume our discussion of the Markov source coding problem of linear coding over finite rings. This is a special scenario of Problem 2 with $s = 1$, g being an identity function and $\mathscr{M} = \left\{ X_1^{(n)} \right\} = \left\{ Y^{(n)} \right\}$ being irreducible Markov. It is known from [10] that the achievable coding rate region for compressing $\mathscr{M}$ is $\{R \in \mathbb{R} | R > H(\mathbf{P}|\pi)\}$ where $\mathbf{P}$ and π are the transition matrix and invariant distribution of $\mathscr{M}$, respectively. Unfortunately, the structures of the encoders used in [10] are unclear (as their Slepian–Wolf correspondences) which limits their application (to Problem 2) as we will see in later sections. This motivates the study of encoders with explicit algebraic structures. We will examine the achievability of linear encoder over a finite ring for this special scenario of Problem 2. The significance of this to other more general settings, where s and g are both arbitrary, will be seen in Section V.

Theorem IV.1. *Assume that $s = 1$, $\mathscr{X}_1 = \mathscr{Y}$ is some finite ring $\mathfrak{R}$ and g is an identity function in Problem 2, and additionally $\left\{ X_1^{(n)} \right\} = \left\{ Y^{(n)} \right\}$ is irreducible Markov with transition matrix $\mathbf{P}$ and invariant distribution*

π. We have that

$$R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \right) \right\}, \tag{10}$$

where

$$\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} = \operatorname{diag}\left\{ \left\{ \mathbf{S}_A \right\}_{A \in \mathfrak{R}/\mathfrak{I}} \right\}$$

with $\mathbf{S}_A$ being the stochastic complement of $\mathbf{P}_{A,A}$ in $\mathbf{P}$ and $Y_{\mathfrak{R}/\mathfrak{I}}^{(i)} = X_1^{(i)} + \mathfrak{I}$, is achievable with linear coding over $\mathfrak{R}$. To be more precise, for any $\epsilon > 0$, there is an $N_0 \in \mathbb{N}^+$ such that there exist a linear encoder $\phi : \mathfrak{R}^n \to \mathfrak{R}^k$ and a decoder $\psi : \mathfrak{R}^k \to \mathfrak{R}^n$ for all $n > N_0$ with

$$\Pr\left\{ \psi\left(\phi\left(Y^n \right) \right) \neq Y^n \right\} < \epsilon,$$

provided that

$$k > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{n}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \right) \right\}.$$

Proof: Let

$$R_0 = \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \right) \right\}$$

and, for any $R > R_0$ and $n \in \mathbb{N}^+$, let $k = \left\lfloor \frac{nR}{\log |\mathfrak{R}|} \right\rfloor$. Obviously, for any $0 \neq \mathfrak{I} \leq_l \mathfrak{R}$ and $\frac{\log |\mathfrak{I}|}{\log |\mathfrak{R}|} \frac{R - R_0}{2} > \eta > 0$, if $n > \frac{2 \log |\mathfrak{I}|}{\eta}$, then

$$\begin{aligned} \left(\frac{\log |\mathfrak{I}|}{\log |\mathfrak{R}|} R_0 - \frac{k}{n} \log |\mathfrak{I}| \right) < & \left(\frac{\log |\mathfrak{I}|}{\log |\mathfrak{R}|} R - 2\eta - \frac{k}{n} \log |\mathfrak{I}| \right) \\ \leq & \frac{\log |\mathfrak{I}|}{n} - 2\eta \\ < & -3\eta/2. \end{aligned}$$

Let $N_0' = \max_{o \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{2 \log |\mathfrak{I}|}{\eta}$. We have that

$$\begin{aligned} & \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \right) \right\} + \eta - \frac{k}{n} \log |\mathfrak{I}| \\ \leq & \frac{\log |\mathfrak{I}|}{\log |\mathfrak{R}|} R_0 + \eta - \frac{k}{n} \log |\mathfrak{I}| \\ = & -\eta/2 \end{aligned} \tag{11}$$

for all $n > N_0'$. The following proves that R is achievable with linear coding over $\mathfrak{R}$.

1) Encoding:

   Choose some $n \in \mathbb{N}^+$ and generate a $k \times n$ matrix $\mathbf{A}$ over $\mathfrak{R}$ uniformly at random (independently choose each entry of $\mathbf{A}$ from $\mathfrak{R}$ uniformly at random). Let the encoder be the linear mapping

   $$\phi : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}, \forall\, \mathbf{x} \in \mathfrak{R}^n.$$

   We note that the coding rate is $\frac{1}{n} \log |\phi(\mathfrak{R}^n)| \leq \frac{1}{n} \log \left| \mathfrak{R}^k \right| = \frac{\log |\mathfrak{R}|}{n} \left\lfloor \frac{nR}{\log |\mathfrak{R}|} \right\rfloor \leq R.$

2) Decoding:

   Choose an $\epsilon > 0$. Assume that $\mathbf{z} \in \mathfrak{R}^k$ is the observation, the decoder claims that $\mathbf{x} \in \mathfrak{R}^n$ is the original data sequence encoded, if and only if

a) $\mathbf{x} \in \mathcal{S}_\epsilon(n, \mathbf{P})$; and

b) $\forall\, \mathbf{x}' \in \mathcal{S}_\epsilon(n, \mathbf{P})$, if $\mathbf{x}' \neq \mathbf{x}$, then $\phi(\mathbf{x}') \neq \mathbf{z}$. In other words, the decoder ψ maps $\mathbf{z}$ to $\mathbf{x}$.

3) Error:

Assume that $\mathbf{X} \in \mathfrak{R}^n$ is the original data sequence generated. An error occurs if and only if

E_1 $\mathbf{X} \notin \mathcal{S}_\epsilon(n, \mathbf{P})$; or

E_2 There exists $\mathbf{x}' \in \mathcal{S}_\epsilon(n, \mathbf{P})$ such that $\phi(\mathbf{x}') = \phi(\mathbf{X})$.

4) Error Probability:

We claim that there exist $N_0 \in \mathbb{N}^+$ and $\epsilon_0 > 0$, if $n > N_0$ and $\epsilon_0 > \epsilon > 0$, then $\Pr\{\psi(\phi(\mathbf{X})) \neq \mathbf{X}\} = \Pr\{E_1 \cup E_2\} < \eta$. First of all, by the AEP of Supremus typicality (Proposition III.4), there exist $N_0'' \in \mathbb{N}^+$ and $\epsilon_0'' > 0$ such that $\Pr\{E_1\} < \eta/2$ if $n > N_0''$ and $\epsilon_0'' > \epsilon > 0$. Secondly, let E_1^c be the complement of E_1. We have

$$\begin{aligned}
&\Pr\{ E_2 | E_1^c\} \\
&= \sum_{\mathbf{x}' \in \mathcal{S}_\epsilon \setminus \{\mathbf{X}\}} \Pr\{ \phi(\mathbf{x}') = \phi(\mathbf{X}) | E_1^c\} \\
&\leq \sum_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \sum_{\mathbf{x}' \in S_\epsilon(\mathbf{X}, \mathfrak{I}) \setminus \{\mathbf{X}\}} \Pr\{ \phi(\mathbf{x}') = \phi(\mathbf{X}) | E_1^c\} && (12) \\
&< \sum_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \exp_2\left[n(r_{\mathfrak{R}/\mathfrak{I}} + \eta)\right] |\mathfrak{I}|^{-k} && (13) \\
&< \left(2^{|\mathfrak{R}|} - 2\right) \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \exp_2\left[n\left(r_{\mathfrak{R}/\mathfrak{I}} + \eta - \frac{k}{n}\log|\mathfrak{I}|\right)\right] && (14) \\
&< \left(2^{|\mathfrak{R}|} - 2\right) \exp_2(-n\eta/2), && (15)
\end{aligned}$$

where $r_{\mathfrak{R}/\mathfrak{I}} = \min\left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H(\mathbf{P}|\pi) - \lim_{m\to\infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right)\right\}$,

(12) follows from the fact that $\mathcal{S}_\epsilon(n, \mathbf{P}) = \bigcup_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} S_\epsilon(\mathbf{X}, \mathfrak{I})$;

(13) is from the typicality lemmas, Lemma III.5 and Lemma III.6, and Lemma II.12, and it is required that ϵ is smaller than some $\epsilon_0''' > 0$ and n is larger than some $N_0''' \in \mathbb{N}^+$;

(14) is due to the fact that the number of non-trivial left ideals of $\mathfrak{R}$ is bounded by $2^{|\mathfrak{R}|} - 2$;

(15) is from (11), and it is required that $n > N_0'$.

Let $N_0 = \max\left\{N_0', N_0'', N_0''', \left\lceil \frac{2}{\eta} \log\left[\frac{2}{\eta}\left(2^{|\mathfrak{R}|} - 2\right)\right]\right\rceil\right\}$ and $\epsilon_0 = \min\{\epsilon_0'', \epsilon_0'''\}$. We have that

$$\Pr\{ E_2 | E_1^c\} < \eta/2 \text{ and } \Pr\{E_1^c\} < \eta/2$$

if $n > N_0$ and $\epsilon_0 > \epsilon > 0$. Hence, $\Pr\{E_1 \cup E_2\} = \Pr\{ E_2 | E_1^c\} + \Pr\{E_1^c\} < \eta$.

By 1) – 4), the theorem is established. ■

Remark 8. From the proof of Theorem IV.1, it is seen that we use the Supremus typicality encoding-decoding technique, in contrast to the traditional (weakly/strongly) typical sequence argument. Technically speaking, if one uses a traditional (weakly/strongly) typical sequence argument, Lemma III.5 will not apply. Consequently, the traditional argument will only achieve the inner bound

$$R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|} \left[H(\mathbf{P}|\pi) - \lim_{m\to\infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right)\right], \tag{16}$$

of (10). Similarly, the inner bound

$$R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} | \pi), \tag{17}$$

is achieved if applying only Lemma III.5 (but not Lemma III.6). Obviously, (10) is the union of these two inner bounds. However, as we have mentioned before, (16) is hard to access in general due to engaging with the entropy rate. Thus, based on (16), it is often hard to draw a optimality conclusion regarding compressing a Markov source as seen below.

Example IV.2. Let $\mathscr{M}$ be an irreducible Markov chain with state space $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ and transition matrix $\mathbf{P} = [p_{i,j}]_{i,j \in \mathbb{Z}_4}$ defined by (9). With simple calculation, (17) says that

$$R > \max\{1.8629, 1.7582\} = H(\mathbf{P}|\pi),$$

where π is the invariant distribution of $\mathscr{M}$, is achievable with linear coding over $\mathbb{Z}_4$. Optimality is attained, i.e. (10) and (17) coincide with the optimal achievable region (cf. [10]). On the contrary, the achievable rate (16) drawn from the traditional typicality argument does not lead to the same optimality conclusion. Because there is no efficient method to evaluate the entropy rate in (16), since neither the initial distribution is known, nor $\mathbf{P} = c_1 \mathbf{U} + (1 - c_1)\mathbf{1}$ for any $\mathbf{U}$ and c_1.

Generally speaking, $\mathscr{X}$ or $\mathscr{Y}$ is not necessarily associated with any algebraic structure. In order to apply the linear encoder, we usually assume that $\mathscr{Y}$ in Problem 2 is mapped into a finite ring $\mathfrak{R}$ of order at least $|\mathscr{Y}|$ by some injection $\Phi : \mathscr{Y} \to \mathfrak{R}$ and denote the set of all possible injections by $\mathcal{I}(\mathscr{Y}, \mathfrak{R})$.

Theorem IV.3. *Assume that $s = 1$, g is an identity function and $\left\{X_1^{(n)}\right\} = \left\{Y^{(n)}\right\}$ is irreducible Markov with transition matrix $\mathbf{P}$ and invariant distribution π in Problem 2. For a finite ring $\mathfrak{R}$ of order at least $|\mathscr{Y}|$ and $\forall\ \Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R})$, let*

$$r_\Phi = \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\Phi,\mathfrak{I}} | \pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) \right\},$$

where

$$\mathbf{S}_{\Phi,\mathfrak{I}} = \mathrm{diag}\left\{ \left\{ \mathbf{S}_{\Phi^{-1}(A)} \right\}_{A \in \mathfrak{R}/\mathfrak{I}} \right\}$$

with $\mathbf{S}_{\Phi^{-1}(A)}$ being the stochastic complement of $\mathbf{P}_{\Phi^{-1}(A),\Phi^{-1}(A)}$ in $\mathbf{P}$ and $Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} = \Phi\left(X_1^{(m)}\right) + \mathfrak{I}$, and define

$$\mathcal{R}_\Phi = \left\{ R \in \mathbb{R} | R > r_\Phi \right\}.$$

We have that

$$\bigcup_{\Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R})} \mathcal{R}_\Phi \tag{18}$$

is achievable with linear coding over $\mathfrak{R}$.

Proof: The result follows immediately from Theorem IV.1 by a timesharing argument. ■

Remark 9. In Theorem IV.3, assume that $\mathscr{Y}$ is some finite ring itself, and let τ be the identity mapping in $\mathcal{I}(\mathscr{Y}, \mathscr{Y})$. It could happen that $\mathcal{R}_\tau \subsetneq \mathcal{R}_\Phi$ for some $\Phi \in \mathcal{I}(\mathscr{Y}, \mathscr{Y})$. This implies that region given by (10) could be strictly smaller than (18). Therefore, a "reordering" of elements in the ring $\mathscr{Y}$ is required when seeking for better linear encoders.

Remark 10. By Lemma B.1, if, in Theorem IV.1, $\mathbf{P} = c_1\mathbf{U} + (1-c_1)\mathbf{1}$ with all rows of $\mathbf{U}$ being identical and $0 \leq c_1 \leq 1$, then

$$R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} \middle| Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)} \right) \right\}$$

is achievable with linear coding over $\mathfrak{R}$. Similarly, if $\mathbf{P} = c_1\mathbf{U} + (1-c_1)\mathbf{1}$ in Theorem IV.3, then, for all $\Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R})$,

$$\mathcal{R}_\Phi = \left\{ R \in \mathbb{R} \middle| R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\Phi,\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} \middle| Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)} \right) \right\} \right\}.$$

Although the achievable regions presented in the above theorems are comprehensive, they depict the optimal one in many situations, i.e. (18) (or (10)) is identical to $H(\mathbf{P}|\pi)$. This has been demonstrated in Example IV.2 above, and more is shown in the following.

Corollary IV.4. *In Theorem IV.1 (or Theorem IV.3), if $\mathfrak{R}$ is a finite field, then*

$$R > H(\mathbf{P}|\pi)$$
$$(\textit{or } \mathcal{R}_\Phi = \left\{ R \in \mathbb{R} \,|\, R > H(\mathbf{P}|\pi) \right\}, \forall\, \Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R}),)$$

is achievable with linear coding over $\mathfrak{R}$.

Proof: If $\mathfrak{R}$ is a finite field, then $\mathfrak{R}$ is the only non-trivial left ideal of itself. The statement follows, since $\mathbf{S}_{\mathfrak{R}/\mathfrak{R}} = \mathbf{P}$ ($\mathbf{S}_{\Phi,\mathfrak{R}} = \mathbf{P}$) and $H\left(Y_{\mathfrak{R}/\mathfrak{R}}^{(m)}\right) = 0$ for all feasible m. ■

Corollary IV.5. *In Theorem IV.3, if $\mathbf{P}$ describes an i.i.d. process, i.e. the row vectors of $\mathbf{P}$ are identical to $\pi = [p_j]_{j \in \mathscr{Y}}$, then*

$$\mathcal{R}_\Phi = \left\{ R \in \mathbb{R} \middle| R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \left[H(\pi) - H(\pi_{\Phi,\mathfrak{I}}) \right] \right\}, \forall\, \Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R}),$$

where $\pi_{\Phi,\mathfrak{I}} = \left[\sum_{j \in \Phi^{-1}(A)} p_j \right]_{A \in \mathfrak{R}/\mathfrak{I}}$*, is achievable with linear coding over $\mathfrak{R}$. In particular, if*

1) *$\mathfrak{R}$ is a field with $|\mathfrak{R}| \geq |\mathscr{Y}|$; or*
2) *$\mathfrak{R}$, with $|\mathfrak{R}| \geq |\mathscr{Y}|$, contains one and only one proper non-trivial left ideal $\mathfrak{I}_0$ and $|\mathfrak{I}_0| = \sqrt{|\mathfrak{R}|}$; or*
3) *$\mathfrak{R}$ is a product ring of several rings satisfying condition 1) or 2),*

then $\bigcup_{\Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R})} \mathcal{R}_\Phi = \left\{ R \in \mathbb{R} \,|\, R > H(\pi) \right\}.$

Proof: The first half of the statement follows from Theorem IV.3 by direct calculation. The second half is from [38]. ■

Remark 11. Concrete examples of the finite ring from Corollary IV.5 includes, but are not limited to:

1) $\mathbb{Z}_p$, where $p \geq |\mathscr{Y}|$ is a prime, as a finite field;
2) $\mathbb{Z}_{p^2}$ and $M_{L,p} = \left\{ \begin{bmatrix} x & 0 \\ y & x \end{bmatrix} \middle| x, y \in \mathbb{Z}_p \right\}$, where $p \geq \sqrt{|\mathscr{Y}|}$ is a prime;

3) $M_{L,p_1} \times \mathbb{Z}_{p_2}$, where $p_1 \geq |\mathscr{Y}|$ and $p_2 \geq |\mathscr{Y}|$ are primes.

Since there always exists a prime p with $p^2 \geq |\mathscr{Y}|$ in Theorem IV.3, Corollary IV.5 guarantees that there always exist optimal linear encoders over some non-field ring, say $\mathbb{Z}_{p^2}$ or $M_{L,p}$, if the source is i.i.d. [38].

Corollary IV.5 can be generalized to the multiple sources scenario in a memoryless setting (see [1], [38]). More precisely, the Slepian–Wolf region is always achieved with linear coding over some non-field ring. Unfortunately, it is neither proved nor denied that a corresponding existence conclusion for the (single or multivariate [39]) Markov source(s) scenario holds. Nevertheless, Example IV.2, Corollary IV.5 and [38] do affirmatively support such an assertion to their own extents.

Even if it is unproved that linear coding over non-field ring is optimal for the scenario of Problem 2 considered in this section, it will be seen in later sections that linear coding over non-field ring strictly outperforms its field counterpart in other settings of this problem.

V. Source Coding for Encoding Markovian Functions

We now move on to a more general setting of Problem 2, where both s and g are arbitrary. Generally speaking, $\mathcal{R}[g]$ is unknown when g is not an identity function (e.g. the binary sum), and it is larger (strictly in many cases) than the Slepian–Wolf region. However, not much is known for the case of sources with memory. Let

$$\mathcal{R}_s = \left\{ [R_1, R_2, \cdots, R_s] \in \mathbb{R}^s \middle| \sum_{t \in T} R_t > \lim_{n \to \infty} \frac{1}{n} \Big[H\left(X^{(n)}, X^{(n-1)}, \cdots, X^{(1)}\right) - H\left(X_{T^c}^{(n)}, X_{T^c}^{(n-1)}, \cdots, X_{T^c}^{(1)}\right) \Big], \emptyset \neq T \subseteq \mathcal{S} \right\}^{8}, \tag{19}$$

where $T^c = \mathcal{S} \setminus T$ and $X_T^{(n)}$ is the random variable array $\prod_{t \in T} X_t^{(n)}$. By [10], if the process

$$\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots$$

is *jointly ergodic*[9] (stationary ergodic), then $\mathcal{R}_s = \mathcal{R}[g]$ for an identity function g. Naturally, $\mathcal{R}_s$ is an inner bound for $\mathcal{R}[g]$ in the case of an arbitrary g. But $\mathcal{R}_s$ is not always tight (optimal), i.e. $\mathcal{R}_s \subsetneq \mathcal{R}[g]$, as we will demonstrate later in Example V.1. Even for the special scenario of correlated i.i.d. sources, i.e. $\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots$ is i.i.d., $\mathcal{R}_s$, which is then the Slepian–Wolf region, is not tight (optimal) in general. Unfortunately, little is mentioned in the existing literature regarding the situation that $\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots$ is not memoryless, neither for the case that $\cdots, Y^{(1)}, Y^{(2)}, \cdots, Y^{(n)}, \cdots$ is homogeneous Markovian (which does not necessarily imply that $\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots$ is jointly ergodic or homogeneous Markov (see Example V.3)).

We begin with briefing the reader on our main idea with Example V.1 in the following. This example shows that the achievable coding rate region for computing a linear function g of s variables is likely to be strictly larger than $\mathcal{R}_s$ in the setting of sources with memory.

[8] Assume the limits exist.

[9] Jointly ergodic defined by Cover [10] is equivalent to stationary ergodic, a condition supporting the Shannon–McMillan–Breiman Theorem. Stationary ergodic is a special case of a.m.s. ergodic [27]. The later is a sufficient and necessary condition for the Point-wise Ergodic Theorem to hold [27, Theorem 1]. The Shannon–McMillan–Breiman Theorem holds under this universal condition as well [27].

Example V.1. Consider three sources S_1, S_2 and S_3 generating random data $X_1^{(i)}$, $X_2^{(i)}$ and $X_3^{(i)}$ (at time $i \in \mathbb{N}^+$) whose sample spaces are all $\mathscr{X}_1 = \mathscr{X}_2 = \mathscr{X}_3 = \{0,1\} \subsetneq \mathbb{Z}_4$, respectively. Let $g : \mathscr{X}_1 \times \mathscr{X}_2 \times \mathscr{X}_3 \to \mathbb{Z}_4$ be defined as

$$g : (x_1, x_2, x_3) \mapsto x_1 + 2x_2 + 3x_3, \tag{20}$$

and assume that $\left\{X^{(n)}\right\}$, where $X^{(i)} = \left(X_1^{(i)}, X_2^{(i)}, X_3^{(i)}\right)$, forms a Markov chain with transition matrix

	(0, 0, 0)	(0, 0, 1)	(0, 1, 0)	(0, 1, 1)	(1, 0, 0)	(1, 0, 1)	(1, 1, 0)	(1, 1, 1)
(0, 0, 0)	.1397	.4060	.0097	.0097	.0097	.0097	.4060	.0097
(0, 0, 1)	.0097	.5360	.0097	.0097	.0097	.0097	.4060	.0097
(0, 1, 0)	.0097	.4060	.1397	.0097	.0097	.0097	.4060	.0097
(0, 1, 1)	.0097	.4060	.0097	.1397	.0097	.0097	.4060	.0097
(1, 0, 0)	.0097	.4060	.0097	.0097	.1397	.0097	.4060	.0097
(1, 0, 1)	.0097	.4060	.0097	.0097	.0097	.1397	.4060	.0097
(1, 1, 0)	.0097	.4060	.0097	.0097	.0097	.0097	.5360	.0097
(1, 1, 1)	.0097	.4060	.0097	.0097	.0097	.0097	.4060	.1397

In order to recover g at the decoder, one solution is to apply Cover's method [10] to first decode the original data and then compute g. This results in an achievable region

$$\mathcal{R}_3 = \left\{ [R_1, R_2, R_3] \in \mathbb{R}^3 \middle| \sum_{t \in T} R_t > \lim_{m \to \infty} \left[H\left(X_1^{(m)}, X_2^{(m)}, X_3^{(m)} \middle| X_1^{(m-1)}, X_2^{(m-1)}, X_3^{(m-1)} \right) \right.\right.$$
$$\left.\left. - H\left(X_{T^c}^{(m)} \middle| X_{T^c}^{(m-1)} \right) \right], \emptyset \neq T \subseteq \{1,2,3\} \right\}.$$

However, $\mathcal{R}_3$ is not optimal, i.e. coding rates beyond this region can be achieved. Observe that $\left\{Y^{(n)}\right\}$, where $Y^{(i)} = g\left(X^{(i)}\right)$, is an irreducible Markovian with transition matrix

	0	3	2	1
0	.1493	.8120	.0193	.0193
3	.0193	.9420	.0193	.0193
2	.0193	.8120	.1493	.0193
1	.0193	.8120	.0193	.1493

(21)

By Theorem IV.1, for any $\epsilon > 0$, there is an $N_0 \in \mathbb{N}^+$, such that for all $n > N_0$ there exist a linear encoder $\phi : \mathbb{Z}_4^n \to \mathbb{Z}_4^k$ and a decoder $\psi : \mathbb{Z}_4^k \to \mathbb{Z}_4^n$, such that $\Pr\left\{\psi\left(\phi\left(Y^n\right)\right) \neq Y^n\right\} < \epsilon$, where $Y^n = \left[Y^{(1)}, Y^{(2)}, \cdots, Y^{(n)}\right]$, as long as

$$k > \frac{n}{2} \times \max\{0.3664, 0.3226\} = 0.1832n.$$

Further notice that

$$\phi\left(Y^n\right) = \vec{g}\left(Z_1^k, Z_2^k, Z_3^k\right),$$

where $Z_t^k = \phi\left(X_t^n\right)$ $(t = 1, 2, 3)$ and $\vec{g}\left(Z_1^k, Z_2^k, Z_3^k\right) = \begin{bmatrix} g\left(Z_1^{(1)}, Z_2^{(1)}, Z_3^{(1)}\right) \\ g\left(Z_1^{(2)}, Z_2^{(2)}, Z_3^{(2)}\right) \\ \vdots \\ g\left(Z_1^{(k)}, Z_2^{(k)}, Z_3^{(k)}\right) \end{bmatrix}$, since g is linear. Thus, another approach [10] is to use ϕ as encoder for each source. Upon observing Z_1^k, Z_2^k and Z_3^k, the decoder claims that $\psi\left(\vec{g}\left(Z_1^k, Z_2^k, Z_3^k\right)\right)$ is the desired data $\vec{g}\left(X_1^n, X_2^n, X_3^n\right)$. Obviously

$$\begin{aligned} &\Pr\left\{\psi\left(\vec{g}\left[\phi\left(X_1^n\right), \phi\left(X_2^n\right), \phi\left(X_3^n\right)\right]\right) \neq Y^n\right\} \\ =& \Pr\left\{\psi\left(\phi\left(Y^n\right)\right) \neq Y^n\right\} < \epsilon, \end{aligned}$$

as long as $k > 0.1832n$. As a consequence, the region

$$\mathcal{R}_{\mathbb{Z}_4} = \left\{ [r, r, r] \in \mathbb{R}^3 \,\middle|\, r > \frac{2k}{n} = 0.4422 \right\} \tag{22}$$

is achieved. Since

$$0.4422 + 0.4422 + 0.4422 < \lim_{m \to \infty} H\left(X_1^{(m)}, X_2^{(m)}, X_3^{(m)} \,\middle|\, X_1^{(m-1)}, X_2^{(m-1)}, X_3^{(m-1)}\right) = 1.4236,$$

we have that $\mathcal{R}_{\mathbb{Z}_4}$ is larger than $\mathcal{R}_3$. In conclusion, $\mathcal{R}_3$ is suboptimal for computing g.

Compared to the one stated in Example V.1, the native Problem 2 is too arbitrary in the sense that even the stochastic property of the sources is unspecified. In order to obtain meaningful conclusions, we will further assume that either condition (c0) or condition (c1) holds. It is easy to see that Example V.1 falls in the category of (c0) which is in fact a special subclass of (c1). One practical interpretation of the mechanism (c0) illustrates is as the following:

> The datum generated at time $n+1$ ($n \in \mathbb{N}^+$) by each source inclines to be the same as the one generated at time n. However, due to some "interference" introduced by the system, the generated data can vary based on a distribution $[u_x]_{x \in \mathscr{X}}$ (a unitary vector). The weights of the two impacts are quantified by $1 - c_1$ and c_1, respectively.

As a special case of (c0), if $c_1 = 1$, then the generated data sequence forms a correlated i.i.d. process. On the other hand, the scene described by (c1) is much broader as mentioned. For instance, g can be a sum of two sources with non-ergodic stochastic behavior, while the sum itself is Markovian. A very interesting realization of such a phenomenon is given later in Example V.3 after the following theorem.

Theorem V.2. *In Problem 2, assume that g satisfies (c1), and let $\mathbf{P}$ and π be the transition matrix and invariant distribution of $\left\{ Z^{(n)} = \sum_{t \in \mathcal{S}} k_t\left(X_t^{(n)}\right) \right\}$, respectively. We have*

$$\mathcal{R} = \left\{ [R, R, \cdots, R] \in \mathbb{R}^s \middle| R > R_0 \right\} \subseteq \mathcal{R}[g],$$

[10] The idea of this approach is first introduced by Körner and Marton [5] for computing the modulo-two sum of two correlated i.i.d. sources. The idea is to present the function as a sum of linear terms (in other words, an Abelian group function). It has been known for long from the work of Körner–Marton [5] and Han–Kobayashi [13]. Csiszár mentioned that these phenomena observed by Körner–Marton partially motivate his investigation on linear encoders over finite fields [3]. In [13, Proof of Theorem 2], an arbitrary function is presented as a sum, although the word "group" is never mentioned. The same idea is then used in [16] which refers to it as "embedding" (into an Abelian group function).

where

$$R_0 = \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \min \left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m \to \infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)} \right) \right\},$$

$\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} = \mathrm{diag}\left\{ \{\mathbf{S}_A\}_{A \in \mathfrak{R}/\mathfrak{I}} \right\}$ *with* $\mathbf{S}_A$ *being the stochastic complement of* $\mathbf{P}_{A,A}$ *in* $\mathbf{P}$ *and* $Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} = Z^{(m)} + \mathfrak{I}$*. Moreover, if* $\mathfrak{R}$ *is a field, then*

$$\mathcal{R} = \left\{ [R, R, \cdots, R] \in \mathbb{R}^s \, | R > H(\mathbf{P}|\pi) \right\}. \tag{23}$$

Proof: Let $Z^n = (Z^{(1)}, Z^{(2)}, \cdots, Z^{(n)})$. By Theorem IV.1, for any $\epsilon > 0$, there exists an $N_0 \in \mathbb{N}^+$ and for all $n > N_0$, there exist an linear encoder $\phi_0 : \mathfrak{R}^n \to \mathfrak{R}^k$ and a decoder $\psi_0 : \mathfrak{R}^k \to \mathfrak{R}^n$ such that

$$\Pr \left\{ \psi_0 \left(\phi_0 \left(Z^n \right) \right) \neq Z^n \right\} < \epsilon,$$

provided that $k > \frac{nR_0}{\log |\mathfrak{R}|}$. Choose $\phi_t = \phi_0 \circ \vec{k}_t$ $(t \in \mathcal{S})$ as the encoder for the tth sources and $\psi = \psi_0 \circ \gamma$, where $\gamma : \mathfrak{R}^s \to \mathfrak{R}$ is defined as $\gamma(x_1, x_2, \cdots, x_s) = \sum_{t \in \mathcal{S}} x_t$, as the decoder. We have that

$$\begin{aligned} & \Pr \left\{ \psi \left(\phi_1 \left(X_1^n \right), \phi_2 \left(X_2^n \right), \cdots, \phi_s \left(X_s^n \right) \right) \neq Z^n \right\} \\ = & \Pr \left\{ \psi_0 \left(\gamma \left(\phi_0 \left(\vec{k}_t \left(X_t^n \right) \right) \right) \right) \neq Z^n \right\} \\ = & \Pr \left\{ \psi_0 \left(\phi_0 \left(\gamma \left(\vec{k}_t \left(X_t^n \right) \right) \right) \right) \neq Z^n \right\} \\ = & \Pr \left\{ \psi_0 \left(\phi_0 \left(Z^n \right) \right) \neq Z^n \right\} < \epsilon. \end{aligned}$$

Therefore, $[r, r, \cdots r] \in \mathbb{R}^s$, where $r = \frac{k \log |\mathfrak{R}|}{n} > R_0$, is achievable for computing g. As a conclusion, $\mathcal{R} \subseteq \mathcal{R}[g]$. If furthermore $\mathfrak{R}$ is a field, then $\mathfrak{R}$ is the only non-trivial left ideal of itself. (23) follows. ■

The following example illustrates a specific instance of (c1) that is not included in (c0). This example is very interesting because it illustrates a scenario where the sources are not jointly ergodic (stationary ergodic) nor a.m.s. ergodic. Thus, [10] does not apply. Yet, Theorem V.2 still provides a solution.

Example V.3. Define $\mathbf{P}_\alpha$ and $\mathbf{P}_\beta$ to be

	(0, 0, 0)	(0, 0, 1)	(0, 1, 0)	(0, 1, 1)	(1, 0, 1)	(1, 1, 0)	(1, 1, 1)	(1, 0, 0)
(0, 0, 0)	.2597	.2093	.2713	.2597	0	0	0	0
(0, 0, 1)	.1208	.0872	.6711	.1208	0	0	0	0
(0, 1, 0)	.0184	.2627	.4101	.3088	0	0	0	0
(0, 1, 1)	.0985	.1823	.2315	.4877	0	0	0	0
(1, 0, 1)	.12985	.10465	.13565	.12985	.12985	.10465	.13565	.12985
(1, 1, 0)	.0604	.0436	.33555	.0604	.0604	.0436	.33555	.0604
(1, 1, 1)	.0092	.13135	.20505	.1544	.0092	.13135	.20505	.1544
(1, 0, 0)	.04925	.09115	.11575	.24385	.04925	.09115	.11575	.24385

and

	(0, 0, 0)	(0, 0, 1)	(0, 1, 0)	(0, 1, 1)	(1, 0, 1)	(1, 1, 0)	(1, 1, 1)	(1, 0, 0)
(0, 0, 0)	0	0	0	0	.2597	.2093	.2713	.2597
(0, 0, 1)	0	0	0	0	.1208	.0872	.6711	.1208
(0, 1, 0)	0	0	0	0	.0184	.2627	.4101	.3088
(0, 1, 1)	0	0	0	0	.0985	.1823	.2315	.4877
(1, 0, 1)	.2597	.2093	.2713	.2597	0	0	0	0
(1, 1, 0)	.1208	.0872	.6711	.1208	0	0	0	0
(1, 1, 1)	.0184	.2627	.4101	.3088	0	0	0	0
(1, 0, 0)	.0985	.1823	.2315	.4877	0	0	0	0

,

respectively. Let $\mathscr{M} = \left\{X^{(n)}\right\}$ be a non-homogeneous Markov chain whose transition matrix from time n to time $n+1$ is

$$\mathbf{P}^{(n)} = \begin{cases} \mathbf{P}_\alpha; & n \text{ is even,} \\ \mathbf{P}_\beta; & \text{otherwise.} \end{cases}$$

Consider Example V.1 by replacing the original homogeneous Markov chain $\left\{X^{(n)}\right\}$ with $\mathscr{M}$ defined above. It is easy to verify that there exists no invariant distribution π' such that $\pi'\mathbf{P}^{(n)} = \pi'$ for all feasible n. This implies that $\mathscr{M}$ is not jointly ergodic (stationary ergodic), nor a.m.s. ergodic. Otherwise, $\mathscr{M}$ will always possess an invariant distribution induced from the *stationary mean measure* of the a.m.s. dynamical system describing $\mathscr{M}$ [40, Theorem 7.1 and Theorem 8.1]. As a consequence, [10] does not apply. However, g Markovian although $\mathscr{M}$ is not even homogeneous. In exact terms, $\left\{g\left(X^{(n)}\right)\right\}$ is homogeneous irreducible Markovian with transition matrix $\mathbf{P}$ given by (9). Consequently, Theorem V.2 offers a solution which achieves

$$\mathcal{R} = \{[r, r, r] | r > H(\mathbf{P}|\pi) = 1.8629\},$$

where π is the unique eigenvector satisfying $\pi\mathbf{P} = \pi$. Once again, the optimal coding rate $H(\mathbf{P}|\pi)$ for compressing $\left\{g\left(X^{(n)}\right)\right\}$ is derived from the Supremus typicality argument, other than the classical (strongly/weakly) typicality argument.

For an arbitrary g, Lemma II.15 promises that there always exist some finite ring $\mathfrak{R}$ and functions $k_t : \mathscr{X}_t \rightarrow$

$\mathfrak{R}$ $(t \in \mathcal{S})$ and $h : \mathfrak{R} \to \mathscr{Y}$ such that

$$g = h\left(\sum_{t \in S} k_t\right).$$

However, $k = \sum_{t \in S} k_t$ is not necessarily Markovian, unless the process $\mathscr{M} = \left\{X^{(n)}\right\}$ is Markov with transition matrix $c_1\mathbf{U} + (1 - c_1)\mathbf{1}$ as stated in (c0). In that case, k is always Markovian so claimed by Lemma B.1.

Corollary V.4. *In Problem 2, assume that $\left\{X^{(n)}\right\}$ forms an irreducible Markov chain with transition matrix $\mathbf{P}_0 = c_1\mathbf{U} + (1 - c_1)\mathbf{1}$, where all rows of $\mathbf{U}$ are identical to some unitary vector and $0 \leq c_1 \leq 1$. Then there exist some finite ring $\mathfrak{R}$ and functions $k_t : \mathscr{X}_t \to \mathfrak{R}$ $(t \in \mathcal{S})$ and $h : \mathfrak{R} \to \mathscr{Y}$ such that*

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t=1}^{s} k_t(x_t)\right) \tag{24}$$

and $\mathscr{M} = \left\{Z^{(n)} = \sum_{t=1}^{s} k_t\left(X_t^{(n)}\right)\right\}$ is irreducible Markov. Furthermore, let π and $\mathbf{P}$ be the invariant distribution and the transition matrix of $\mathscr{M}$, respectively, and define

$$R_0 = \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|} \min\left\{H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}\middle|Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}\right)\right\}$$

where $\mathbf{S}_{\mathfrak{R}/\mathfrak{I}} = \operatorname{diag}\left\{\{\mathbf{S}_A\}_{A \in \mathfrak{R}/\mathfrak{I}}\right\}$ with $\mathbf{S}_A$ being the stochastic complement of $\mathbf{P}_{A,A}$ in $\mathbf{P}$ and $Y_{\mathfrak{R}/\mathfrak{I}}^{(m)} = Z^{(m)} + \mathfrak{I}$. We have that

$$\mathcal{R}_{\mathfrak{R}} = \{[R, R, \cdots, R] \in \mathbb{R}^s | R > R_0\} \subseteq \mathcal{R}[g]. \tag{25}$$

Proof: The existences of k_t's and h are from Lemma II.15, and Lemma B.1 ensures that $\mathscr{M}$ is Markovian. In addition, $\left\{X^{(n)}\right\}$ is irreducible, so is $\mathscr{M}$. Finally,

$$\lim_{m\to\infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) = \lim_{m\to\infty} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}\middle|Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}\right),$$

since $\left\{Y_{\mathfrak{R}/\mathfrak{I}}^{(n)}\right\}$ is Markovian by Lemma B.1. This implies that $\mathcal{R}_{\mathfrak{R}} \subseteq \mathcal{R}[g]$ by Theorem V.2. ■

Remark 12. It is easy to verify that the irreducibility requirement in (c0) is equivalent to that $u_x > 0$ for all $x \in \mathscr{X}$. Besides, if $c_1 = 1$, then (c0) renders to the memoryless scenario, [1, Problem 1]. If this is the case, Corollary V.4 resumes corresponding results of [1, Section VI] (see Corollary V.5).

Remark 13. For the function g in Corollary V.4, it is often the case that there exists more than one finite ring $\mathfrak{R}$ or more than one set of functions k_t's and h satisfying corresponding requirements. For example [1], the polynomial function $x + 2y + 3z \in \mathbb{Z}_4[3]$ admits also the polynomial presentation $\hat{h}\left(x + 2y + 4z\right) \in \mathbb{Z}_5[3]$, where $\hat{h}(u) = \sum_{a \in \mathbb{Z}_5} a\left[1 - (u - a)^4\right] - \left[1 - (u - 4)^4\right] \in \mathbb{Z}_5[1]$. As a conclusion, a better inner bound of $\mathcal{R}[g]$ is

$$\mathcal{R}_s \bigcup \left(\bigcup_{\mathfrak{R}} \bigcup_{\mathscr{P}_{\mathfrak{R}}(g)} \mathcal{R}_{\mathfrak{R}}\right), \tag{26}$$

where $\mathscr{P}_{\mathfrak{R}}(g)$ denotes all the polynomial presentations of format (24) of g over ring $\mathfrak{R}$.

Corollary V.5. *In Corollary V.4, let $\pi = [p_j]_{j \in \mathfrak{R}}$. If $c_1 = 1$, namely, $\left\{X^{(n)}\right\}$ and $\mathscr{M}$ are i.i.d., then*

$$\mathcal{R}_{\mathfrak{R}} = \left\{[R, R, \cdots, R] \in \mathbb{R}^s \middle| R > \max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(\pi) - H(\pi_{\mathfrak{I}})\right]\right\} \subseteq \mathcal{R}[g], \tag{27}$$

where $\pi_{\mathfrak{I}} = \left[\sum_{j \in A} p_j\right]_{A \in \mathfrak{R}/\mathfrak{I}}$.

Remark 14. In Corollary V.5, under many circumstances it may hold that $\max_{0 \neq \mathfrak{I} \leq_l \mathfrak{R}} \frac{\log |\mathfrak{R}|}{\log |\mathfrak{I}|} \left[H(\pi) - H(\pi_{\mathfrak{I}})\right] = H(\pi)$, i.e.

$$\mathcal{R}_{\mathfrak{R}} = \left\{[R, R, \cdots, R] \in \mathbb{R}^s \left| R > H(\pi) \right.\right\}.$$

For example, when $\mathfrak{R}$ is a field. However, $\mathfrak{R}$ being a field is definitely not necessary. For more details, please kindly refer to [1], [7], [38].

Corollary V.6. *In Corollary V.4,* $\mathfrak{R}$ *can always be chosen as a field. Consequently,*

$$\mathcal{R}_{\mathfrak{R}} = \left\{[R, R, \cdots, R] \in \mathbb{R}^s \left| R > H(\mathbf{P}|\pi) \right.\right\} \subseteq \mathcal{R}[g].$$

Remark 15. Although $\mathfrak{R}$ in Corollary V.4 can always be chosen to be a field, the region $\mathcal{R}_{\mathfrak{R}}$ is not necessarily larger than when $\mathfrak{R}$ is chosen as a non-field ring. On the contrary, $\mathcal{R}_{\mathfrak{R}}$ is strictly larger when $\mathfrak{R}$ is a non-field ring than when it is chosen as a field in many case. This is because the induced $\mathbf{P}$, as well as π, varies.

As mentioned, in Theorem V.2, Corollary V.4 and Corollary V.5, there may be more than one choice of such a finite ring $\mathfrak{R}$ satisfying the corresponding requirements. Among those choices, $\mathfrak{R}$ can be either a field or a non-field ring. Surprisingly, it is seen in (infinitely) many examples that using non-field ring outperforms using a field. In many cases, it is proved that the achievable region obtained with linear coding over some non-field ring is strictly larger than any that is achieved with its field counterpart, regardless which field is considered. [1, Example VI.2] has demonstrated this in the setting of correlated i.i.d. sources. In the next section, this will be once again demonstrated in the setting of sources with memory. In addition, other advantages of the non-field ring linear coding technique will be investigated in comparing with the field version.

VI. Advantages: Non-field Rings versus Fields

Clearly, our discussion regarding linear coding is mainly based on general finite rings which can be either fields or non-field rings, each bringing their own advantages. In the setting where g is the identity function in Problem 2, linear coding over finite field is always optimal in the sense of achieving $\mathcal{R}[g]$ if the sources are jointly ergodic [10]. An equivalent conclusive result is not yet proved for linear coding over non-field ring. Nevertheless, it is proved that there always exist more than one (up to isomorphism) non-field rings over which linear coding achieves the Slepian–Wolf region if the sources considered are i.i.d. [38]. Furthermore, many examples, say Example IV.2, show that non-field ring can be equally optimal when considering Markov sources. All in all, there is still no conclusive support that linear coding over field is preferable in terms of achieving the optimal region $\mathcal{R}[g]$ with g being an identity function.

On the contrary, there are many drawbacks of using finite fields compared to using non-field rings (e.g. modulo integer rings):

1) The finite field arithmetic is complicated to implement since the finite field arithmetic usually involves the *polynomial long division algorithm*; and
2) The alphabet size(s) of the encoder(s) is (are) usually larger than required [1], [7], [8]; and

3) In many specific circumstances of Problem 2, linear coding over any finite field is proved to be less optimal than its non-field rings counterpart in terms of achieving larger achievable region (see [1], [8] and Example VI.1); and
4) The characteristic of a finite field has to be a prime. This constraint creates shortages in their polynomial presentations of discrete functions (see Lemma B.3). These shortages confine the performance of the polynomial approach (if restrict to field) and lead to results like Proposition VI.2. On the other hand, The characteristic can be any positive integer for a finite non-field ring; and
5) Field (finite or not) contains no *zero divisor*. This also impares the performance of the polynomial approach (if restrict to field).

Example VI.1. Consider the situation illustrated in Example V.1, one alternative is to treat that $\mathscr{X}_1 = \mathscr{X}_2 = \mathscr{X}_3 = \{0,1\}$ as a subset of finite field $\mathbb{Z}_5$ and the function g can then be presented as

$$g(x_1, x_2, x_3) = \hat{h}(x_1 + 2x_2 + 4x_3),$$

where $\hat{h} : \mathbb{Z}_5 \to \mathbb{Z}_4$ is given by $\hat{h}(z) = \begin{cases} z; & z \neq 4, \\ 3; & z = 4, \end{cases}$ (symbol-wise). By Corollary V.6, linear coding over $\mathbb{Z}_5$ achieves the region

$$\mathcal{R}_{\mathbb{Z}_5} = \left\{ [r, r, r] \in \mathbb{R}^3 \left| r > H\left(\mathbf{P}_{\mathbb{Z}_5} | \pi_{\mathbb{Z}_5}\right) = 0.4623 \right. \right\}.$$

Obviously, $\mathcal{R}_{\mathbb{Z}_5} \subsetneq \mathcal{R}_{\mathbb{Z}_4} \subseteq \mathcal{R}[g]$. In conclusion, using linear coding over field $\mathbb{Z}_5$ is less optimal compared with over non-field ring $\mathbb{Z}_4$. In fact, the region $\mathcal{R}_{\mathbb{F}}$ achieved by linear coding over any finite field $\mathbb{F}$ is always strictly smaller than $\mathcal{R}_{\mathbb{Z}_4}$.

Proposition VI.2. *In Example V.1, $\mathcal{R}_{\mathbb{F}}$, the achievable region achieved with linear coding over any finite field $\mathbb{F}$ in the sense of Corollary V.4, is properly contained in $\mathcal{R}_{\mathbb{Z}_4}$, i.e. $\mathcal{R}_{\mathbb{F}} \subsetneq \mathcal{R}_{\mathbb{Z}_4}$.*

Proof: Assume that

$$g(x_1, x_2, x_3) = h\left(k_1(x_1) + k_2(x_2) + k_3(x_3)\right)$$

with $k_t : \{0,1\} \to \mathbb{F}$ $(1 \leq t \leq 3)$ and $h : \mathbb{F} \to \mathbb{Z}_4$. Let

$$\mathscr{M}_1 = \left\{Y^{(n)}\right\} \text{ with } Y^{(n)} = g\left(X_1^{(n)}, X_2^{(n)}, X_3^{(n)}\right),$$
$$\mathscr{M}_2 = \left\{Z^{(n)}\right\} \text{ with } Z^{(n)} = k_1\left(X_1^{(n)}\right) + k_2\left(X_2^{(n)}\right) + k_3\left(X_3^{(n)}\right),$$

and $\mathbf{P}_l$ and π_l be the transition matrix and the invariant distribution of $\mathscr{M}_l$, respectively, for $l = 1, 2$. By Corollary V.4 (also Corollary V.6), linear coding over $\mathbb{F}$ achieves the region

$$\mathcal{R}_{\mathbb{F}} = \left\{[R, R, \cdots, R] \in \mathbb{R}^s \left| R > H(\mathbf{P}_2|\pi_2) \right. \right\},$$

while linear coding over $\mathbb{Z}_4$ achieves

$$\mathcal{R}_{\mathbb{Z}_4} = \left\{ [R, R, \cdots, R] \in \mathbb{R}^s \left| R > \max_{0 \neq \mathfrak{I} \leq_l \mathbb{Z}_4} \frac{\log |\mathbb{Z}_4|}{\log |\mathfrak{I}|} H(\mathbf{S}_{\mathbb{Z}_4/\mathfrak{I}}|\pi_1) = H(\mathbf{P}_1|\pi_1) \right. \right\}.$$

Moreover,

$$H(\mathbf{P}_1|\pi_1) < H(\mathbf{P}_2|\pi_2)$$

by Lemma B.2 due to Lemma B.3 claims that $h|_{\mathscr{S}}$, where $\mathscr{S} = k_1(\{0,1\}) + k_2(\{0,1\}) + k_3(\{0,1\})$, can never be injective. Therefore, $\mathcal{R}_{\mathbb{F}} \subsetneq \mathcal{R}_{\mathbb{Z}_4}$. ■

Remark 16. There are infinitely many functions like g defined in Example V.1 such that the achievable region obtained with linear coding over any finite field in the sense of Corollary V.4 is strictly suboptimal compared to the one achieved with linear coding over some non-field ring. These functions include $\sum_{t=1}^{s} x_t \in \mathbb{Z}_{2p}[s]$ for any $s \geq 2$ and any prime $p > 2$. One can always find a concrete example in which linear coding over $\mathbb{Z}_{2p}$ dominates. The reason for this is partially because these functions are defined on rings (e.g. $\mathbb{Z}_{2p}$) of non-prime characteristic. However, a finite field must be of prime characteristic, resulting in conclusions like Proposition VI.2.

As a direct consequence of Proposition VI.2, we have

Theorem VI.3. *In the sense of* (26)*, linear coding over finite field is not optimal.*

VII. Conclusions

This paper considers the ring linear coding technique introduced in [1] in the setting of compressing data generated by a single Markov source. An achievability theorem, as a generalization of its field counterpart, is presented. The paper also demonstrates that the compression limit can be reached with linear encoders over non-field rings. However, this property is not yet conclusively proved in general.

On the other hand, a variation of the data compression problem, namely Problem 2 is addressed. We apply the polynomial approach of [14], [18], [1] to the scenarios where sources are with memory. Once again, it is seen that linear coding technique over non-field ring dominates its field counterpart in terms of achieving better coding rates for computing (encoding) some functions. On this regard, we claim that linear coding over finite field is not optimal.

To facilitate our discussions, the concept of Supremus typical sequence and its related asymptotic properties are introduced. These include the AEP and four generalized typicality lemmas. Compared to the traditional version, Supremus typicality allows us to draw more accessible results, while corresponding ones of traditional typicality are often hard to analyze as demonstrated. The new techniques are hopefully helpful also in understanding and investigating other related problems.

Appendix A
Proof of Proposition II.6

1) Let $\Pr\left\{X^{(1)} = x^{(1)}\right\} = c$. By definition,

$$\begin{aligned}
&\Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} \\
=&\Pr\left\{X^{(1)} = x^{(1)}\right\} \prod_{i,j\in\mathscr{X}} p_{i,j}^{N(i,j;\mathbf{x})} \\
=&c\exp_2\left[\sum_{i,j\in\mathscr{X}} N(i,j;\mathbf{x})\log p_{i,j}\right] \\
=&c\exp_2\left[-n\sum_{i,j\in\mathscr{X}} -\frac{N(i;\mathbf{x})}{n}\frac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})}\log p_{i,j}\right] \\
=&c\exp_2\left[-n\sum_{i,j\in\mathscr{X}}\left(p_ip_{i,j} - \frac{N(i;\mathbf{x})}{n}\frac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})}\right)\log p_{i,j} - p_ip_{i,j}\log p_{i,j}\right].
\end{aligned}$$

In addition, there exists a small enough $\epsilon_0 > 0$ and a $N_0 \in \mathbb{N}^+$ such that $\left|\frac{N(i;\mathbf{x})}{n}\frac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})} - p_ip_{i,j}\right| < -\eta \Big/ 2\left|\mathscr{X}\right|^2 \min_{i,j\in\mathscr{X}} \log p_{i,j}$ and $-\frac{\log c}{n} < \eta/2$ for all $\epsilon_0 > \epsilon > 0$ and $n > N_0$. Consequently,

$$\begin{aligned}
&\Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} \\
>&c\exp_2\left[-n\sum_{i,j\in\mathscr{X}}\frac{\eta}{2\left|\mathscr{X}\right|^2\min_{i,j\in\mathscr{X}}\log p_{i,j}}\log p_{i,j} - p_ip_{i,j}\log p_{i,j}\right] \\
\geq&c\exp_2\left[-n\left(\frac{\eta}{2} - \sum_{i,j\in\mathscr{X}} p_ip_{i,j}\log p_{i,j}\right)\right] \\
=&\exp_2\left[-n\left(-\frac{\log c}{n} + \frac{\eta}{2} + H(\mathbf{P}|\pi)\right)\right] \\
>&\exp_2\left[-n\left(\eta + H(\mathbf{P}|\pi)\right)\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\Pr\left\{\left[X^{(1)}, X^{(2)}, \cdots, X^{(n)}\right] = \mathbf{x}\right\} \\
<&c\exp_2\left[-n\sum_{i,j\in\mathscr{X}}\frac{-\eta}{2\left|\mathscr{X}\right|^2\min_{i,j\in\mathscr{X}}\log p_{i,j}}\log p_{i,j} - p_ip_{i,j}\log p_{i,j}\right] \\
\leq&c\exp_2\left[-n\left(-\frac{\eta}{2} - \sum_{i,j\in\mathscr{X}} p_ip_{i,j}\log p_{i,j}\right)\right] \\
\leq&\exp_2\left[-n\left(-\frac{\eta}{2} + H(\mathbf{P}|\pi)\right)\right] \\
<&\exp_2\left[-n\left(-\eta + H(\mathbf{P}|\pi)\right)\right].
\end{aligned}$$

2) By Boole's inequality [41], [42],

$$\begin{aligned}\Pr\left\{\mathbf{X}\notin\mathcal{T}_\epsilon(n,\mathbf{P})\right\} =&\Pr\left\{\left(\bigcup_{i,j\in\mathscr{X}}\left|\frac{N(i,j;\mathbf{X})}{N(i;\mathbf{X})}-p_{i,j}\right|\geq\epsilon\right)\bigcup\left(\bigcup_{i\in\mathscr{X}}\left|\frac{N(i;\mathbf{X})}{n}-p_i\right|\geq\epsilon\right)\right\}\\ &\leq\sum_{i,j\in\mathscr{X}}\Pr\left\{\left|\frac{N(i,j;\mathbf{X})}{N(i;\mathbf{X})}-p_{i,j}\right|\geq\epsilon\middle|E\right\}+\sum_{i\in\mathscr{X}}\Pr\left\{\left|\frac{N(i;\mathbf{X})}{n}-p_i\right|\geq\epsilon\right\},\end{aligned}$$

where $E=\bigcap_{i\in\mathscr{X}}\left\{\left|\frac{N(i;\mathbf{X})}{n}-p_i\right|<\epsilon\right\}$ for all feasible i.

By the Ergodic Theorem of Markov chains [29, Theorem 1.10.2], $\Pr\left\{\left|\frac{N(i;\mathbf{X})}{n}-p_i\right|\geq\epsilon\right\}\to 0$ as $n\to\infty$ for any $\epsilon>0$. Thus, there is an integer N_0', such that for all $n>N_0'$, $\Pr\left\{\left|\frac{N(i;\mathbf{X})}{n}-p_i\right|\geq\epsilon\right\}<\frac{\eta}{2|\mathscr{X}|}$. On the other hand, for $\min_{i\in\mathscr{X}}p_i/2>\epsilon>0$, $N(i;\mathbf{x})\to\infty$ as $n\to\infty$, conditional on E. Therefore, by the Strong Law of Large Numbers [29, Theorem 1.10.1], $\Pr\left\{\left|\frac{N(i,j;\mathbf{X})}{N(i;\mathbf{X})}-p_{i,j}\right|\geq\epsilon\middle|E\right\}\to 0$, $n\to\infty$. Hence, there exists N_0'', for all $n>N_0''$, $\Pr\left\{\left|\frac{N(i,j;\mathbf{X})}{N(i;\mathbf{X})}-p_{i,j}\right|\geq\epsilon\middle|E\right\}<\frac{\eta}{2|\mathscr{X}|^2}$. Let $N_0=\max\{N_0',N_0''\}$ and $\epsilon_0=\min_{i\in\mathscr{X}}p_i/2>0$. We have $\Pr\left\{\mathbf{X}\notin\mathcal{T}_\epsilon(n,\mathbf{P})\right\}<\eta$ for all $\epsilon_0>\epsilon>0$ and $n>N_0$.

3) Finally, let ϵ_0 and N_0 be defined as in 1). $|\mathcal{T}_\epsilon(n,\mathbf{P})|<\exp_2\left[n\left(H(\mathbf{P}|\pi)+\eta\right)\right]$ follows since

$$\begin{aligned}1\geq&\sum_{\mathbf{x}\in\mathcal{T}_\epsilon(n,\mathbf{P})}\Pr\{\mathbf{X}=\mathbf{x}\}\\ &>|\mathcal{T}_\epsilon(n,\mathbf{P})|\exp_2\left[-n\left(H(\mathbf{P}|\pi)+\eta\right)\right],\end{aligned}$$

if $\epsilon_0>\epsilon>0$ and $n>N_0$.

Let ϵ_0 be the smallest one chosen above and N_0 be the biggest one chosen. The statement is proved.

Appendix B
Supporting Lemmas

Lemma B.1. *Let $\left\{X^{(n)}\right\}$ be a Markov chain with countable state space $\mathscr{X}$ and transition matrix $\mathbf{P}_0$. If $\mathbf{P}_0=c_1\mathbf{U}+(1-c_1)\mathbf{1}$, where $\mathbf{U}$ is a matrix all of whose rows are identical to some countably infinite unitary vector and $0\leq c_1\leq 1$, then $\left\{\Gamma\left(X^{(n)}\right)\right\}$ is Markov for all feasible function Γ.*

Proof: Let $Y^{(n)} = \Gamma\left(X^{(n)}\right)$, and assume that $[u_x]_{x \in \mathscr{X}}$ is the first row of $\mathbf{U}$. For any $a, b \in \Gamma(\mathscr{X})$,

$$\begin{aligned}
&\Pr\left\{ Y^{(n+1)} = b \middle| Y^{(n)} = a \right\} \\
=& \sum_{x \in \Gamma^{-1}(a)} \Pr\left\{ X^{(n)} = x, Y^{(n+1)} = b \middle| Y^{(n)} = a \right\} \\
=& \sum_{x \in \Gamma^{-1}(a)} \Pr\left\{ Y^{(n+1)} = b \middle| X^{(n)} = x, Y^{(n)} = a \right\} \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\} \\
=& \sum_{x \in \Gamma^{-1}(a)} \Pr\left\{ Y^{(n+1)} = b \middle| X^{(n)} = x \right\} \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\} \\
=& \begin{cases} \displaystyle\sum_{x \in \Gamma^{-1}(a)} \sum_{x' \in \Gamma^{-1}(b)} c_1 u_{x'} \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\}; & a \neq b \\ \displaystyle\sum_{x \in \Gamma^{-1}(a)} \left[1 - c_1 + \sum_{x' \in \Gamma^{-1}(b)} c_1 u_{x'} \right] \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\}; & a = b \end{cases} \\
=& \begin{cases} \displaystyle c_1 \sum_{x' \in \Gamma^{-1}(b)} u_{x'} \sum_{x \in \Gamma^{-1}(a)} \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\}; & a \neq b \\ \displaystyle\left[1 - c_1 + c_1 \sum_{x' \in \Gamma^{-1}(b)} u_{x'} \right] \sum_{x \in \Gamma^{-1}(a)} \Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a \right\}; & a = b \end{cases} \\
=& \begin{cases} \displaystyle c_1 \sum_{x' \in \Gamma^{-1}(b)} u_{x'}; & a \neq b \\ \displaystyle 1 - c_1 + c_1 \sum_{x' \in \Gamma^{-1}(b)} u_{x'}; & a = b \end{cases} \\
=& \sum_{x' \in \Gamma^{-1}(b)} \Pr\left\{ X^{(n+1)} = x' \middle| X^{(n)} = x \right\} \left(\forall\, x \in \Gamma^{-1}(a) \right) \\
=& \sum_{x' \in \Gamma^{-1}(b)} \Pr\left\{ X^{(n+1)} = x' \middle| X^{(n)} = x \right\} \Pr\left\{ Y^{(n)} = a \middle| Y^{(n)} = a, Y^{(n-1)}, \cdots \right\} \left(\forall\, x \in \Gamma^{-1}(a) \right) \\
=& \sum_{x \in \Gamma^{-1}(a)} \sum_{x' \in \Gamma^{-1}(b)} \Pr\left\{ X^{(n+1)} = x' \middle| X^{(n)} = x, Y^{(n)} = a, Y^{(n-1)}, \cdots \right\} \\
&\Pr\left\{ X^{(n)} = x \middle| Y^{(n)} = a, Y^{(n-1)}, \cdots \right\} \\
=& \sum_{x \in \Gamma^{-1}(a)} \sum_{x' \in \Gamma^{-1}(b)} \Pr\left\{ X^{(n+1)} = x', X^{(n)} = x \middle| Y^{(n)} = a, Y^{(n-1)}, \cdots \right\} \\
=& \Pr\left\{ Y^{(n+1)} = b \middle| Y^{(n)} = a, Y^{(n-1)}, \cdots \right\}
\end{aligned}$$

Therefore, $\left\{ \Gamma\left(X^{(n)} \right) \right\}$ is Markov. ■

Remark 17. Lemma B.1 is enlightened by [11, Theorem 3]. However, $\left\{ X^{(n)} \right\}$ in this lemma is not necessarily stationary or finite-state.

Lemma B.2. *Let $\mathscr{X}$ be a countable set, $\pi = [p(z)]_{z \in \mathscr{X}}$ and $\mathbf{P} = [p(z_1, z_2)]_{z_1, z_2 \in \mathscr{X}}$ be a non-negative unitary vector and a stochastic matrix, respectively. For any function $h : \mathscr{X} \to \mathscr{Y}$, if for all $y_1, y_2 \in \mathscr{Y}$*

$$\frac{p(z_1, y_2)}{p(z_1)} = c_{y_1, y_2}, \forall\, z_1 \in h^{-1}(y_1), \tag{28}$$

where c_{y_1,y_2} is a constant, then

$$H\left(h\left(Z^{(2)}\right)\middle|h\left(Z^{(1)}\right)\right) \leq H(\mathbf{P}|\pi), \tag{29}$$

where $\left(Z^{(1)}, Z^{(2)}\right) \sim \pi\mathbf{P}$. Moreover, (29) *holds with equality if and only if*

$$p(z_1, h(z_2)) = p(z_1, z_2), \forall\, z_1, z_2 \in \mathscr{Z} \text{ with } p(z_1, z_2) > 0. \tag{30}$$

Proof: By definition,

$$\begin{aligned}
&H\left(h\left(Z^{(2)}\right)\middle|h\left(Z^{(1)}\right)\right)\\
=&-\sum_{y_1,y_2\in\mathscr{Y}} p(y_1,y_2)\log\frac{p(y_1,y_2)}{p(y_1)}\\
=&-\sum_{y_1,y_2\in\mathscr{Y}}\sum_{z_1\in h^{-1}(y_1)} p(z_1,y_2)\log\left(\sum_{z_1'\in h^{-1}(y_1)} p(z_1',y_2)\middle/\sum_{z_1''\in h^{-1}(y_1)} p(z_1'')\right)\\
\overset{(a)}{=}&-\sum_{y_1,y_2\in\mathscr{Y}}\sum_{z_1\in h^{-1}(y_1)} p(z_1,y_2)\log\frac{p(z_1,y_2)}{p(z_1)}\\
=&-\sum_{y_1,y_2\in\mathscr{Y}}\sum_{\substack{z_2\in h^{-1}(y_2),\\ z_1\in h^{-1}(y_1)}} p(z_1,z_2)\log\frac{\sum_{z_2'\in h^{-1}(y_2)} p(z_1,z_2')}{p(z_1)}\\
\overset{(b)}{\leq}&-\sum_{y_1,y_2\in\mathscr{Y}}\sum_{\substack{z_2\in h^{-1}(y_2),\\ z_1\in h^{-1}(y_1)}} p(z_1,z_2)\log\frac{p(z_1,z_2)}{p(z_1)}\\
=&-\sum_{z_1,z_2\in\mathscr{Z}} p(z_1,z_2)\log\frac{p(z_1,z_2)}{p(z_1)}\\
=&H(\mathbf{P}|\pi),
\end{aligned}$$

where (a) is from (28). In addition, equality holds, i.e. (b) holds with equality, if and only if (30) is satisfied. ■

Remark 18. $\mathbf{P}$ in the above lemma can be interpreted as the transition matrix of some Markov process. However, π is not necessarily the corresponding invariant distribution. It is also not necessary that such a Markov process is irreducible. In the meantime, (29) can be seen as a "data processing inequality". In addition, (28) is sufficient but not necessary for (29), even though it is sufficient and necessary for (a) in the above proof.

Lemma B.3. *For g given by* (20) *and any finite field $\mathbb{F}$, if there exist functions $k_t : \{0,1\} \to \mathbb{F}$ and $h : \mathbb{F} \to \mathbb{Z}_4$, such that*

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t=1}^{s} k_t(x_t)\right),$$

then $h|_{\mathscr{S}}$, where $\mathscr{S} = k_1(\{0,1\}) + k_2(\{0,1\}) + k_3(\{0,1\})$, is not injective.

Proof: Suppose otherwise, i.e. $h|_{\mathscr{S}}$ is injective. Let $h' : h(\mathscr{S}) \to \mathscr{S}$ be the inverse mapping of $h : \mathscr{S} \to$

$h\left(\mathscr{S}\right)$. Obviously, h' is bijective. By (20), we have

$$\begin{aligned} & h'\left[g(1,0,0)\right] = k_1(1) + k_2(0) + k_3(0) \\ =& h'\left[g(0,1,1)\right] = k_1(0) + k_2(1) + k_3(1) \\ \neq& h'\left[g(1,1,0)\right] = k_1(1) + k_2(1) + k_3(0) \\ =& h'\left[g(0,0,1)\right] = k_1(0) + k_2(0) + k_3(1). \end{aligned}$$

Let $\tau = h'\left[g(1,0,0)\right] - h'\left[g(1,1,0)\right] = h'\left[g(0,1,1)\right] - h'\left[g(0,0,1)\right] \in \mathbb{F}$. We have that

$$\begin{aligned} & \tau = k_2(0) - k_2(1) = k_2(1) - k_2(0) = -\tau \\ \Longrightarrow & \tau + \tau = 0. \end{aligned} \tag{31}$$

(31) implies that either $\tau = 0$ or Char($\mathbb{F}$) $= 2$ by [1, Proposition II.6]. Noticeable that $k_2(0) \neq k_2(1)$, i.e. $\tau \neq 0$, by the definition of g. Thus, Char($\mathbb{F}$) $= 2$. Let $\rho = k_3(0) - k_3(1)$. Obviously, $\rho \neq 0$ by the definition of g, and $\rho + \rho = 0$ since Char($\mathbb{F}$) $= 2$. Consequently,

$$\begin{aligned} h'\left[g(0,0,0)\right] =& k_1(0) + k_2(0) + k_3(0) \\ =& k_1(0) + k_2(0) + k_3(1) + \rho \\ =& h'\left[g(0,0,1)\right] + \rho \\ =& h'\left[g(1,1,0)\right] + \rho \\ =& k_1(1) + k_2(1) + k_3(0) + \rho \\ =& k_1(1) + k_2(1) + k_3(1) + \rho + \rho \\ =& h'\left[g(1,1,1)\right]. \end{aligned}$$

Therefore, $g(0,0,0) = g(1,1,1)$ since h' is bijective. This is absurd! ■

Appendix C

Typicality Lemmas of Supremus Typical Sequences

Given a set $\mathscr{X}$, a *partition* $\coprod_{k \in \mathscr{K}} A_k$ of $\mathscr{X}$ is a disjoint union of $\mathscr{X}$, i.e. $A_{k'} \cap A_{k''} \neq \emptyset \Leftrightarrow k' = k''$, $\bigcup_{k \in \mathscr{K}} A_k = \mathscr{X}$ and A_k's are not empty. Obviously, $\coprod_{A \in \mathfrak{R}/\mathfrak{I}} A$ is a partition of a ring $\mathfrak{R}$ given the left (right) ideal $\mathfrak{I}$.

Lemma C.1. *Given an irreducible Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}$ with finite state space $\mathscr{X}$, transition matrix $\mathbf{P}$ and invariant distribution $\pi = \left[p_j\right]_{j \in \mathscr{X}}$. Let $\coprod_{k=1}^{m} A_k$ be any partition of $\mathscr{X}$. For any $\eta > 0$, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\ \epsilon_0 > \epsilon > 0$, $\forall\ n > N_0$ and $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P})$,*

$$|S_\epsilon(\mathbf{x})| < \exp_2 \left\{ n \left[\sum_{k=1}^{m} \sum_{j \in A_k} p_j H(\mathbf{S}_k | \pi_k) + \eta \right] \right\} \tag{32}$$

$$= \exp_2 \left\{ n \left[H(\mathbf{S} | \pi) + \eta \right] \right\} \tag{33}$$

where

$$S_\epsilon(\mathbf{x}) = \left\{ \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P}) \middle| y^{(l)} \in A_k \Leftrightarrow x^{(l)} \in A_k, \forall\, 1 \leq l \leq n, \forall\, 1 \leq k \leq m \right\},$$

$\mathbf{S}_k$ *is the stochastic complement of* $\mathbf{P}_{A_k, A_k}$ *in* $\mathbf{P}$*,* $\pi_k = \dfrac{[p_i]_{i \in A_k}}{\sum_{j \in A_k} p_j}$ *is the invariant distribution of* $\mathbf{S}_k$ *and*

$$\mathbf{S} = \operatorname{diag}\left\{ \{\mathbf{S}_k\}_{1 \leq k \leq m} \right\}.$$

Proof: Let

$$\mathbf{x}_{A_k} = \left[x^{(n_1)}, x^{(n_2)}, x^{(n_{m_k})}\right]$$

be the subsequence of $\mathbf{x}$ formed by all those $x^{(l)}$'s belong to A_k in the original ordering. Obviously, $\sum_{k=1}^{m} m_k = n$ and $\left| \dfrac{m_k}{n} - \sum_{j \in A_k} p_j \right| < |A_k| \epsilon + \dfrac{1}{n}$. For any $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in S_\epsilon(\mathbf{x})$,

$$\mathbf{y}_{A_k} = \left[y^{(n_1)}, y^{(n_2)}, y^{(n_{m_k})}\right] \in A_k^{m_k}$$

is a strongly Markov ϵ-typical sequence of length m_k with respect to $\mathbf{S}_k$ by Proposition III.2, since $\mathbf{y}$ is Supremus ϵ-typical. Additionally, by Proposition II.6, there exist $\epsilon_k > 0$ and positive integer M_k such that the number of strongly Markov ϵ-typical sequences of length m_k is upper bounded by $\exp_2 \{m_k [H(\mathbf{S}_k|\pi_k) + \eta/2]\}$ if $0 < \epsilon < \epsilon_k$ and $m_k > M_k$. Therefore, if $0 < \epsilon < \min_{1 \leq k \leq m} \epsilon_k$, $n > M = \max_{1 \leq k \leq m} \left\{ \dfrac{1 + M_k}{\left|\sum_{j \in A_k} p_j - |A_k| \epsilon\right|} \right\}$ (this guarantees that $m_k > M_k$ for all $1 \leq k \leq m$), then

$$\begin{aligned} |S_\epsilon(\mathbf{x})| \leq& \exp_2 \left\{ \sum_{k=1}^{m} m_k [H(\mathbf{S}_k|\pi_k) + \eta/2] \right\} \\ =& \exp_2 \left\{ n \left[\sum_{k=1}^{m} \frac{m_k}{n} H(\mathbf{S}_k|\pi_k) + \eta/2 \right] \right\}. \end{aligned}$$

Furthermore, choose $0 < \epsilon_0 \leq \min_{1 \leq k \leq m} \epsilon_k$ and $N_0 \geq M$ such that $\dfrac{m_k}{n} < \sum_{j \in A_k} p_j + \dfrac{\eta}{2 \sum_{k=1}^{m} H(\mathbf{S}_k|\pi_k)}$ for all $0 < \epsilon < \epsilon_0$ and $n > N_0$ and $1 \leq k \leq m$, we have

$$|S_\epsilon(\mathbf{x})| < \exp_2 \left\{ n \left[\sum_{k=1}^{m} \sum_{j \in A_k} p_j H(\mathbf{S}_k|\pi_k) + \eta \right] \right\},$$

(32) is established. Direct calculation yields (33). ■

By definition, $S_\epsilon(\mathbf{x})$ in Lemma C.1 contains Supremus ϵ-typical sequences whose have the same sequential pattern as $\mathbf{x}$ regarding the partition $\coprod_{k=1}^{m} A_k$. Similarly, let $T_\epsilon(\mathbf{x})$ be the set of strongly Markov ϵ-typical sequences with the same sequential pattern as $\mathbf{x}$ regarding the partition $\coprod_{k=1}^{m} A_k$, namely

$$T_\epsilon(\mathbf{x}) = \left\{ \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in \mathcal{T}_\epsilon(n, \mathbf{P}) \middle| y^{(l)} \in A_k \Leftrightarrow x^{(l)} \in A_k, \forall\, 1 \leq l \leq n, \forall\, 1 \leq k \leq m \right\}.$$

We have that

Lemma C.2. *In Lemma C.1, define* $\Gamma(x) = l \Leftrightarrow x \in A_l$*. We have that*

$$|S_\epsilon(\mathbf{x})| \leq |T_\epsilon(\mathbf{x})| < \exp_2 \left\{ n \left[H(\mathbf{P}|\pi) - \lim_{w \to \infty} \frac{1}{w} H\left(Y^{(w)}, Y^{(w-1)}, \cdots, Y^{(1)}\right) + \eta \right] \right\},$$

where $Y^{(w)} = \Gamma\left(X^{(w)}\right)$.

Proof: $|S_\epsilon(\mathbf{x})| \leq |T_\epsilon(\mathbf{x})|$ is trivial. Let

$$\overline{\mathbf{y}} = \left[\Gamma\left(x^{(1)}\right), \Gamma\left(x^{(2)}\right), \cdots, \Gamma\left(x^{(n)}\right)\right].$$

By definition,

$$\left[\Gamma\left(y^{(1)}\right), \Gamma\left(y^{(2)}\right), \cdots, \Gamma\left(y^{(n)}\right)\right] = \overline{\mathbf{y}},$$

for any $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in S_\epsilon(\mathbf{x})$. $\mathbf{y}$ is *jointly typical* [10] with $\overline{\mathbf{y}}$ with respect to the process

$$\cdots, \begin{pmatrix} X^{(1)} \\ Y^{(1)} \end{pmatrix}, \begin{pmatrix} X^{(2)} \\ Y^{(2)} \end{pmatrix}, \cdots, \begin{pmatrix} X^{(n)} \\ Y^{(n)} \end{pmatrix}, \cdots$$

Therefore, there exist $\epsilon_0 > 0$ and $N_0 \in \mathbb{N}^+$, such that, $\forall\, \epsilon_0 > \epsilon > 0$ and $\forall\, n > N_0$,

$$\begin{aligned} |S_\epsilon(\mathbf{x})| <& \exp_2\left\{n\left[\lim_{w\to\infty}\frac{1}{w}H\left(X^{(w)}, X^{(w-1)}, \cdots, X^{(1)}\right)\right.\right. \\ &\left.\left. - \lim_{w\to\infty}\frac{1}{w}H\left(Y^{(w)}, Y^{(w-1)}, \cdots, Y^{(1)}\right) + \eta\right]\right\} \\ =& \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - \lim_{w\to\infty}\frac{1}{w}H\left(Y^{(w)}, Y^{(w-1)}, \cdots, Y^{(1)}\right) + \eta\right]\right\}, \end{aligned}$$

where the equality follows from the fact that $\lim_{w\to\infty}\frac{1}{w}H\left(X^{(w)}, X^{(w-1)}, \cdots, X^{(1)}\right) = H\left(\mathbf{P}\,|\pi\,\right)$ since $\mathscr{M}$ is irreducible Markov. ■

Remark 19. Given a left ideal $\mathfrak{I}$ of a finite ring $\mathfrak{R}$, $\mathfrak{R}/\mathfrak{I}$ gives raise to a partition of $\mathfrak{R}$. Let $\mathscr{X} = \mathfrak{R}$, $m = |\mathfrak{R}/\mathfrak{I}|$ and A_k $(1 \leq k \leq m)$ be an element (which is a set) of $\mathfrak{R}/\mathfrak{I}$. One has Lemma III.5 and Lemma III.6 proved immediately. In fact, Lemma C.1 and Lemma C.2 can be easily tailored to corresponding versions regarding other algebraic structures, e.g. group, rng[11], vector space, module, algebra and etc, in a similar fashion.

[11] A ring without multiplicative identity.

References

[1] S. Huang and M. Skoglund, *On Linear Coding over Finite Rings and Applications to Computing*, KTH Royal Institute of Technology, October 2012. [Online]. Available: http://people.kth.se/∼sheng11

[2] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

[3] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, Jul. 1982.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[5] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.

[6] I. Csiszár, Private Communication, Jul. 2013.

[7] S. Huang and M. Skoglund, "On achievability of linear source coding over finite rings," in *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2013, pp. 1984–1988.

[8] ——, "On existence of optimal linear encoders over non-field rings for data compression with application to computing," in *IEEE Information Theory Workshop*, September 2013.

[9] C. D. Meyer, "Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems," *SIAM Rev.*, vol. 31, no. 2, pp. 240–272, Jun. 1989. [Online]. Available: http://dx.doi.org/10.1137/1031050

[10] T. M. Cover, "A proof of the data compression theorem of slepian and wolf for ergodic sources," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 226–228, Mar. 1975.

[11] C. J. Burke and M. Rosenblatt, "A Markovian function of a Markov chain," *The Annals of Mathematical Statistics*, vol. 29, no. 4, pp. 1112–1122, Dec. 1958. [Online]. Available: http://www.jstor.org/stable/2236949

[12] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 396–411, May 1983.

[13] T. S. Han and K. Kobayashi, "A dichotomy of functions f(x, y) of correlated sources (x, y) from the viewpoint of the achievable rate region," *IEEE Transactions on Information Theory*, vol. 33, no. 1, pp. 69–76, Jan. 1987.

[14] S. Huang and M. Skoglund, "Polynomials and computing functions of correlated sources," in *IEEE International Symposium on Information Theory*, Jul. 2012, pp. 771–775.

[15] H. Yamamoto, "Wyner–ziv theory for a general function of the correlated sources," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 803–807, 1982.

[16] D. Krithivasan and S. Pradhan, "Distributed source coding using Abelian group codes: A new achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1495–1519, 2011.

[17] M. Sefidgaran and A. Tchamkerten, "Computing a function of correlated sources: A rate region," in *IEEE International Symposium on Information Theory*, Aug. 2011, pp. 1856–1860.

[18] S. Huang and M. Skoglund, "Computing polynomial functions of correlated sources: Inner bounds," in *International Symposium on Information Theory and its Applications*, Oct. 2012, pp. 160–164.

[19] ——, "Linear source coding over rings and applications," in *IEEE Swedish Communication Technologies Workshop*, Oct. 2012, pp. 1–6.

[20] G. Como and F. Fagnani, "The capacity of finite Abelian group codes over symmetric memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2037–2054, May 2009.

[21] A. Orlitsky and R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, Mar. 2001.

[22] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498 –3516, Oct. 2007.

[23] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, "Network coding for computing: Cut-set bounds," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 1015–1030, 2011.

[24] V. Lalitha, N. Prakash, K. Vinodh, P. Kumar, and S. Pradhan, "Linear coding schemes for the distributed computation of subspaces," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 678–690, 2013.

[25] L. D. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 431–438, Jul. 1981.

[26] I. Csiszár, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[27] R. M. Gray and J. C. Kieffer, "Asymptotically mean stationary measures," *The Annals of Probability*, vol. 8, no. 5, pp. 962–973, Oct. 1980. [Online]. Available: http://www.jstor.org/stable/2242939

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, Jul. 2006.

[29] J. R. Norris, *Markov Chains*. Cambridge University Press, Jul. 1998.

[30] L. Breuer and D. Baum, *An Introduction to Queueing Theory: and Matrix-Analytic Methods*, 2005th ed. Springer, Dec. 2005.

[31] D. S. Dummit and R. M. Foote, *Abstract Algebra*, 3rd ed. Wiley, 2003.

[32] J. J. Rotman, *Advanced Modern Algebra*, 2nd ed. American Mathematical Society, Aug. 2010.

[33] T. W. Hungerford, *Algebra (Graduate Texts in Mathematics)*. Springer, Dec. 1980.

[34] T.-Y. Lam, *A First Course in Noncommutative Rings*, 2nd ed. Springer, Jun. 2001.

[35] R. C. Buck, "Nomographic functions are nowhere dense," *Proceedings of the American Mathematical Society*, vol. 85, no. 2, pp. 195–199, Jun. 1982. [Online]. Available: http://www.jstor.org/stable/2044280

[36] S. Huang and M. Skoglund, *Induced Transformations of Recurrent A.M.S. Dynamical Systems*, KTH Royal Institute of Technology, October 2013. [Online]. Available: http://people.kth.se/∼sheng11

[37] J. Aaronson, *An Introduction to Infinite Ergodic Theory*. Providence, R.I.: American Mathematical Society, 1997.

[38] S. Huang and M. Skoglund, *On Existence of Optimal Linear Encoders over Non-field Rings for Data Compression*, KTH Royal Institute of Technology, December 2012. [Online]. Available: http://people.kth.se/∼sheng11

[39] E. Fung, W. K. Ching, S. Chu, M. Ng, and W. Zang, "Multivariate Markov chain models," in *2002 IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, Oct. 2002.

[40] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 2nd ed. Springer, Aug. 2009.

[41] G. Boole, *An investigation of the laws of thought on which are founded, the mathematical theories of logic and probabilities*. [S.l.]: Watchmaker, 2010.

[42] M. Fréchet, "Généralisation du théorème des probabilités totales," *Fundamenta Mathematicae*, vol. 25, no. 1, pp. 379–387, 1935. [Online]. Available: https://eudml.org/doc/212798