Coding for Computing Irreducible Markovian Functions of Sources with Memory

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

Abstract

One open problem in source coding is to characterize the limits of representing losslessly a non-identity discrete function of the data encoded independently by the encoders of several correlated sources with memory. This paper investigates this problem under Markovian conditions, namely either the sources or the functions considered are Markovian. We propose using linear mappings over finite rings as encoders. If the function considered admits certain polynomial structure, the linear encoders can make use of this structure to establish "implicit collaboration" and boost the performance. In fact, this approach universally applies to any scenario (arbitrary function) because any discrete function admits a polynomial presentation of required format.

There are several useful discoveries in the paper. The first says that linear encoder over non-field ring can be equally optimal for compressing data generated by an irreducible Markov source. Secondly, regarding the previous function-encoding problem, there are infinitely many circumstances where linear encoder over non-field ring strictly outperforms its field counterpart. To be more precise, it is seen that the set of coding rates achieved by linear encoder over certain non-field rings is strictly larger than the one achieved by the field version, regardless which finite field is considered. Therefore, in this sense, linear coding over finite field is not optimal. In addition, for certain scenarios where the sources do not possess the ergodic property, our ring approach is still able to offer a solution.

Index Terms

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

I. Introduction

This paper considers the problem of encoding a *discrete function* of correlated sources with memory:

Problem 2 (Source Coding for Computing a Function of Sources with or without Memory)**.** 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 $\forall$ $\epsilon > 0$,

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.

$\exists\ N_0 \in \mathbb{N}^+$, 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

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

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 touple $\mathbf{R} \in \mathbb{R}^s$ is said to be *achievable* for computing g (or simply achievable) if and only 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 and only 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., so is

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

Related work for this special scenario includes: [2], [3] which considers the case that g is an identity function; [4], [5] where g is the binary sum; [6], [7] for conditions under which that $\mathcal{R}[g]$ is strictly larger than the Slepian–Wolf region; [8], [9], [10], [11], [1] for an arbitrary discrete function g. Generally speaking, $\mathcal{R}[g]$ is unknown in cases where g is not an identity function, and it is larger (strictly in many cases) than the Slepian–Wolf region.

Furthermore, much less is known in the case of sources with memory. Let

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

where $T^c = \mathcal{S} \setminus T$ and $X_T^{(n)}$ is the random variable array $\prod_{t \in T} X_t^{(n)}$. By [12], if the process

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

is *jointly ergodic* (see [12] for definition), then $\mathcal{R}_s = \mathcal{R}[g]$ for an identity function g. Naturally, $\mathcal{R}_s$ is an inner bound of $\mathcal{R}[g]$ for an arbitrary g. However, $\mathcal{R}_s$ 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.

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

[1] Assume the limits exist.

is i.i.d., $\mathcal{R}_s$, which is then the Slepian–Wolf region, is not tight (optimal) in general as mentioned before. Unfortunately, little is mentioned in existing literature regarding the case

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

is not memoryless, neither for the case that

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

is Markovian (which does not necessary imply that $\cdots, X^{(1)}, X^{(2)}, \cdots, X^{(n)}, \cdots$ is jointly ergodic or Markov).

This paper focuses on Problem 2 in the sense that some additional Markovian constraints are imposed since the original scenario is too general. We 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

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t\in\mathcal{S}} k_t(x_t)\right), \tag{2}$$

such that $\left\{\sum_{t\in\mathcal{S}} k_t\left(X_t^{(n)}\right)\right\}_{-\infty}^{\infty}$ is irreducible[2] Markovian[3].

By Lemma II.15 and Lemma C.1, (c1) includes a very interesting scenario:

(c0) g is arbitrary, while

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

forms an irreducible Markov chain with transition matrix

$$\mathbf{P}_0 = c_1\mathbf{U} + (1 - c_1)\mathbf{1}, \tag{3}$$

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 $0 \leq c_1 \leq 1$.

If, as a special case, $c_1 = 1$, then Problem 2 renders to [1, Problem 1], since

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

becomes i.i.d.. Actually, (c0) is very interesting because of the fact:

A stationary finite-state Markov chain

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

admits a transition matrix of the form (3), if and only if

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

is Markovian for all feasible mappings Γ [13, Theorem 3].

We will explain the mechanism that (c0) illustrates when the discussion comes. Here we would like to point out that (c1) is a rather general assumption. It even includes some scenario that

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

[2] Irreducibility of a Markov chain / process is sometimes (implicitly) assumed in some literature.

[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.

does not possess the ergodic property (see Example V.3). Therefore, [12] does not apply and (1) does not present an achievable region. However, it is sometimes possible to classify such a scenario as a special case of (c1), to which a solution is provided in this paper (see Section V).

This paper aims at developing similar results as [1] based on this new setting. To be more precise, we will first prove an achievability theorem for source coding with linear encoder over finite ring for compressing a single finite-state irreducible Markov source. This generalizes the corresponding theorem regarding linear encoder over field. Making use of the linear coding technique introduced by this achievability theorem, we then address Problem 2 of computing g regarding each of the previous conditions, (c0) and (c1). Inner bounds of $\mathcal{R}[g]$ are presented. It is demonstrated that the achievable regions given by these inner bounds are beyond (1). Even more interestingly, our method (for computing some g) even works for cases in which

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

does not possess the ergodic property. 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.

Apart from classic information theoretical techniques, the key mathematical tools involved are the uncoupling-coupling technique and the concept of stochastic complement of finite-state Markov processes (see [14] for more details). With the aid of these tools, we will introduce the concept of Supremus typical sequences (Definition III.1) and prove related asymptotic properties (Proposition III.2) and typicality lemmas (Appendix D). These serve as the foundation of our arguments thereafter.

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. Some Notation

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

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

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

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

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. $\sum_{i\in\mathscr{X}} p_i = 1$, and $[p_{i,j}]_{i\in\mathscr{X},j\in\mathscr{Y}}$ be non-negative and $\sum_{j\in\mathscr{Y}} p_{i,j} = 1$ (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

Definition II.1. A (discrete) *Markov chain* is defined to be a discrete stochastic process $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with *state space* $\mathscr{X}$ such that, $\forall\ n \in \mathbb{N}^+$,

$$\Pr\left\{X^{(n+1)}\middle| X^{(n)}, X^{(n-1)}, \cdots, X^{(1)}\right\} = \Pr\left\{X^{(n+1)}\middle| X^{(n)}\right\}.$$

$\mathscr{M}$ is said to be *finite-state* if $\mathscr{X}$ is finite.

Definition II.2. A Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ is said to be *homogeneous* (*time homogeneous*) if and only if

$$\Pr\left\{X^{(n+1)}\middle| X^{(n)}\right\} = \Pr\left\{X^{(2)}\middle| X^{(1)}\right\}, \forall\ n \in \mathbb{N}^+.$$

If not specified, we assume finite-state and homogeneous of all Markov chains considered throughout this paper. However, they are not necessarily *stationary* [15, pp. 71], or their *initial distribution* is unknown.

Definition II.3. Given a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with a countable 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.4. A state j of a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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.5 (Theorem 1.7.7 of [16])**.** *An irreducible Markov chain $\mathscr{M}$ with a countable 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.6 (Theorem 2.31 of [17])**.** *A finite-state irreducible Markov chain is positive recurrent.*

Clearly, all irreducible Markov chains considered in this paper admit a unique invariant distribution, since they are assumed to be simultaneously finite-state and homogeneous (unless otherwise specified).

Definition II.7 (Strong Markov Typicality)**.** Let $\mathscr{M} = \left\{ X^{(n)} \right\}_{-\infty}^{\infty}$ 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

$$\begin{cases} \left| \dfrac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})} - p_{i,j} \right| < \epsilon; \\ \left| \dfrac{N(i;\mathbf{x})}{n} - p_i \right| < \epsilon, \end{cases} \quad \forall\, i,j \in \mathscr{X}, \tag{4}$$

$$\text{or} \begin{cases} \displaystyle\sum_{i,j \in \mathscr{X}} \left| \dfrac{N(i,j;\mathbf{x})}{N(i;\mathbf{x})} - p_{i,j} \right| < \epsilon; \\ \displaystyle\sum_{i \in \mathscr{X}} \left| \dfrac{N(i;\mathbf{x})}{n} - p_i \right| < \epsilon, \end{cases} \tag{5}$$

where $N(i,j;\mathbf{x})$ is the occurrences of sub-sequence $[i,j]$ in $\mathbf{x}$ and $N(i;\mathbf{x}) = \sum_{j \in \mathscr{X}} N(i,j;\mathbf{x})$. 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.

Remark 1. (4) and (5) is equivalent (in illustrating the asymptotic behavior of $\mathscr{M}$) to

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

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

for some fixed finite constant c, respectively.

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.8 (AEP of Strongly Markov Typicality[4])**.** *Let $\mathscr{M} = \left\{ X^{(n)} \right\}_{-\infty}^{\infty}$ 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,*

[4] Similar statements in many literature assume that the Markov chain is stationary. It is easy to generalize to irreducible Markov chain. To be rigorous, we include a proof in Appendix A.

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) $|\mathcal{T}_\epsilon(n, \mathbf{P})| < \exp_2\left[n\left(H(\mathbf{P}|\pi) + \eta\right)\right]$.

Proof: See Appendix A. ■

Remark 2. For a strongly Markov ϵ-typical sequence $(\mathbf{x}, \mathbf{y})^t \in \mathscr{X}^n \times \mathscr{Y}^n$, it is not necessary that $\mathbf{x}$ or $\mathbf{y}$ is strongly Markov ϵ-typical. As a matter of fact, given an irreducible Markov chain $\left\{\left(X^{(n)}, Y^{(n)}\right)^t\right\}_{-\infty}^{\infty}$, stochastic processes $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ or $\left\{Y^{(n)}\right\}_{-\infty}^{\infty}$ is not necessary Markov.

C. Rings, Ideals and Linear Mappings

Definition II.9. The touple $[\mathfrak{R}, +, \cdot]$ 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.9, 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.10 (c.f. [18]). 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 3. 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 [19, Proposition 2.137].

Definition II.11. 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 $\forall\ r \in \mathfrak{R}$, $x \cdot r \in \mathfrak{I}$,

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 $\mathfrak{I} \leq_l \mathfrak{R}$ or $\mathfrak{I} \leq_r \mathfrak{R}$, then $\mathfrak{R}$ is divided into disjoint *cosets* which are 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 forms a *quotient group*, denoted by $\mathfrak{R}/\mathfrak{I}$ (see [19, Ch. 1.6 and Ch. 2.9] for more details).

Definition II.12. A mapping $f : \mathfrak{R}^n \to \mathfrak{R}^m$ given as:

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

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.

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

$$g(x_1, x_2, \cdots, x_k) = \sum_{j=0}^m a_j x_1^{m_{1j}} x_2^{m_{2j}} \cdots x_k^{m_{kj}}, \tag{6}$$

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 4. *Polynomial* and polynomial function are sometimes only defined over a commutative ring [19]. It is a very delicate matter to define them over a non-commutative ring [20], [21], due to the fact that $x_1 x_2$ and $x_2 x_1$ can become different objects. We choose to define "polynomial functions" with formula (6) because those functions are within the scope of this paper's interest.

Lemma II.14. *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 $\nu : \mathscr{Y} \to \mathfrak{R}$.

Proof: Let p be a prime such that $p^m \geq \max\left\{|\mathscr{Y}|, |\mathscr{X}_i| \,|1 \leq i \leq k\right\}$ for some integer m, and choose $\mathfrak{R}$ to be a finite field of order p^m. By [22, Lemma 7.40], the number of polynomial functions in $\mathfrak{R}[k]$ is $p^{mp^{mk}}$. Moreover, the number of distinct functions with domain $\mathfrak{R}^k$ and codomain $\mathfrak{R}$ is also $|\mathfrak{R}|^{\left|\mathfrak{R}^k\right|} = p^{mp^{mk}}$. Hence, any function $g : \mathfrak{R}^k \to \mathfrak{R}$ is a polynomial function.

In the meanwhile, any injections $\mu_i : \mathscr{X}_i \to \mathfrak{R}$ ($1 \leq i \leq k$) and $\nu : \mathscr{Y} \to \mathfrak{R}$ give rise to a function

$$\hat{g} = \nu \circ g\left(\mu_1', \mu_2', \cdots, \mu_k'\right) : \mathfrak{R}^k \to \mathfrak{R},$$

where μ_i' is the inverse mapping of $\mu_i : \mathscr{X}_i \to \mu_i(\mathscr{X}_i)$. Since $\hat{g}$ must be a polynomial function as shown, the statement is established. ■

Remark 5. Another proof of Lemma II.14 involving Fermat's little theorem can be found in [7].

The important message conveyed by Lemma II.14 says that any discrete function defined on a finite domain is essentially a *restriction* [7, Definition II.3] of some polynomial function. Therefore, we can restrict the consideration of Problem 2 to all polynomial functions. This polynomial approach[5] offers a very good insight into the general problem. After all, the algebraic structure of a polynomial function is much more clear than 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 [7], [9].

Lemma II.15. *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*

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t=1}^{s} k_t(x_t)\right). \tag{7}$$

Proof: There are several proofs of this lemma. One is provided in appendix B. ■

We often name the polynomial function $\hat{g}$ in Lemma II.14 a *polynomial presentation* of g. This paper mainly focuses on presentations of format (7). Readers are kindly referred to [9] for other interested formats. As a simple demonstration [7], once 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 is of format (7).

III. Stochastic Complement, Reduced Markov Chains and Supremus Typical Sequences

Given a Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 $\mathscr{M}_A = \left\{X^{(T_{A,l})}\right\}_{-\infty}^{\infty}$ is Markov by the strong Markov property [16, 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

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

[5] This polynomial approach is first proposed in [7], [9].

is the transition matrix of $\mathscr{M}_A$ [14, 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. $\pi_A \mathbf{S}_A = \pi_A$ [14, Theorem 2.2]. Since $\mathscr{M}_A$ inherits irreducibility from $\mathscr{M}$ [14, 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* 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 $\mathscr{X}$. The set of all Supremus ϵ-typical sequences with respect to $\mathbf{P}$ in $\mathscr{X}^n$ is denoted $\mathcal{S}_\epsilon(n, \mathbf{P})$ or $\mathcal{S}_\epsilon$ for simplicity.

Proposition III.2 (AEP of Supremus Typicality)**.** *Let $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 6. Motivated by Definition III.1, Proposition III.2 and two related typicality lemmas in Appendix D, one can define the concept of *Supremus type* resembling other classic types [23], e.g. Markov type [24]. We will consider this in our future work for inspecting error exponents of the schemes introduced in this paper.

The following are two typicality lemmas of Supremus typical sequences tailored for our discussions. They are the ring specials of the two given in Appendix D, respectively.

Lemma III.3. *Let $\mathfrak{R}$ be a finite ring, $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 $\forall\ \mathfrak{I} \leq_l \mathfrak{R}$,*

$$|S_\epsilon(\mathbf{x}, \mathfrak{I})| < \exp_2\left\{n\left[\sum_{A \in \mathfrak{R}/\mathfrak{I}} \sum_{j \in A} p_j H(\mathbf{S}_A|\pi_A) + \eta\right]\right\} \tag{8}$$

$$= \exp_2\left\{n\left[H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi) + \eta\right]\right\} \tag{9}$$

where

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

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

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

Proof: 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 $\left|\dfrac{m_A}{n} - \sum_{j\in A} p_j\right| < |A|\,\epsilon + \dfrac{1}{n}$). It is easily seen that $\mathbf{y}_A = \left[y^{(n_1)}, y^{(n_2)}, y^{(n_{m_A})}\right] \in A^{m_A}$ 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.8, 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 $\exp_2\left\{m_A\left[H(\mathbf{S}_A|\pi_A) + \eta/2\right]\right\}$ if $0 < \epsilon < \epsilon_A$ and $m_A > M_A$. Therefore, if $0 < \epsilon < \min_{A\in\mathfrak{R}/\mathfrak{I}} \epsilon_A$, $n > M = \max_{A\in\mathfrak{R}/\mathfrak{I}}\left\{\dfrac{1+M_A}{\left|\sum_{j\in A} p_j - |A|\,\epsilon\right|}\right\}$ (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 $\dfrac{m_A}{n} < \sum_{j\in A} p_j + \dfrac{\eta}{2\sum_{A\in\mathfrak{R}/\mathfrak{I}} H(\mathbf{S}_A|\pi_A)}$ for all $0 < \epsilon < \epsilon_0$ and $n > N_0$ and $A \in \mathfrak{R}/\mathfrak{I}$, we have

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

(8) is established. Direct calculation yields (9). ■

Lemma III.4. *In Lemma III.3,*

$$|S_\epsilon(\mathbf{x}, \mathfrak{I})| < \exp_2\left\{n\left[H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} \frac{1}{m} H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)}, Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)}, \cdots, Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right) + \eta\right]\right\}, \tag{10}$$

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

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

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

$\mathbf{y}$ is *jointly typical* [12] with $\overline{\mathbf{y}}$ 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 $\forall\ n > N_0$,

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

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

Remark 7. In Lemma III.4, 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 $\mathscr{M}' = \left\{Y^{(n)}_{\mathfrak{R}/\mathfrak{I}}\right\}_{-\infty}^{\infty}$ is Markovian by Lemma C.1. As a conclusion,

$$\begin{aligned}|S_\epsilon(\mathbf{x}, \mathfrak{I})| <& \exp_2 \left\{ n \left[H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} H\left(Y^{(m)}_{\mathfrak{R}/\mathfrak{I}} \middle| Y^{(m-1)}_{\mathfrak{R}/\mathfrak{I}}\right) + \eta \right] \right\} \\ =& \exp_2 \left\{ n \left[H\left(\mathbf{P}|\pi\right) - H\left(\mathbf{P}'|\pi'\right) + \eta \right] \right\},\end{aligned} \tag{11}$$

where $\mathbf{P}'$ and π' are the transition matrix and the invariant distribution of $\mathscr{M}'$ that can be easily calculated from $\mathbf{P}$. However, in general $\mathscr{M}'$ is ergodic, but not Markovian. Its *entropy rate* is difficult to obtain.

Remark 8. If $\mathfrak{R}$ in Lemma III.3 is a field, then both (9) and (10) 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.3 is i.i.d., then both (9) and (10) are equivalent to

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

which is a special case of the *generalized conditional typicality lemma* [1, Lemma III.5]. However, it is hard to determine which bound of these two is tighter in general. Nevertheless, (9) is seemingly easier to analyze, while (10) is more complicated for associating with the entropy rate of the ergodic process $\left\{Y^{(n)}_{\mathfrak{R}/\mathfrak{I}}\right\}_{-\infty}^{\infty}$.

Remark 9. Lemma III.3 and Lemma III.4 can be easily generalized to corresponding versions regarding other algebraic structures, e.g. group, rng[6], vector space, module, algebra and etc.

IV. Achievability Theorem of Linear Coding for One Markov Source

Equipped with the foundation laid down by Lemma III.3 and Lemma III.4, we resume our discussion to Problem 2. For the time being, this section only considers a special scenario, namely $s = 1$, g is an identity function and $\mathscr{M} = \left\{X_1^{(n)}\right\}_{-\infty}^{\infty} = \left\{Y^{(n)}\right\}_{-\infty}^{\infty}$ is irreducible Markov in Problem 2. It is known from [12] that the achievable coding rate region for such a scenario 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 [12] are unclear which limits their application (to Problem 2) as we will see in later sections. This motivates our study of encoders with explicit algebraic structures. We will examine the achievability of linear encoder over

[6] A ring without multiplicative identity.

a finite ring for this special scenario of Problem 2. The significance of this to other more general settings of Problem 2, 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 $\left\{X_1^{(n)}\right\}_{-\infty}^{\infty}=\left\{Y^{(n)}\right\}_{-\infty}^{\infty}$ is irreducible Markov with transition matrix $\mathbf{P}$ and invariant distribution π. We have that*

$$R>\max_{0\neq\mathfrak{I}\leq_l\mathfrak{R}}\frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\min\left\{H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi),H\left(\mathbf{P}|\pi\right)-\lim_{m\to\infty}\frac{1}{m}H\left(Y_{\mathfrak{R}/\mathfrak{I}}^{(m)},Y_{\mathfrak{R}/\mathfrak{I}}^{(m-1)},\cdots,Y_{\mathfrak{R}/\mathfrak{I}}^{(1)}\right)\right\},\tag{12}$$

where

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

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

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

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.2. *Assume that $s=1$, g is an identity function and $\left\{X_1^{(n)}\right\}_{-\infty}^{\infty}=\left\{Y^{(n)}\right\}_{-\infty}^{\infty}$ 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

$$\bigcup_{\Phi\in\mathcal{I}(\mathscr{Y},\mathfrak{R})}\mathcal{R}_\Phi\tag{13}$$

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

Proof: The result follows immediately from Theorem IV.1. ■

Remark 10. In Theorem IV.2, 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 (12) can be strictly smaller than (13). Therefore, a "reordering" of elements in the ring $\mathscr{Y}$ is required when seeking for better linear encoders.

Remark 11. By Lemma C.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 $0 \le c_1 \le 1$, then

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

is achievable with linear coding over $\mathfrak{R}$. Similarly, if $\mathbf{P} = c_1\mathbf{U} + (1-c_1)\mathbf{1}$ in Theorem IV.2, then, for all $\Phi \in \mathcal{I}(\mathscr{Y}, \mathfrak{R})$,

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

Proof of Theorem IV.1: Let

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

and, for any $R > R_0$ and $n \in \mathbb{N}^+$, let $k = \left\lfloor \dfrac{nR}{\log|\mathfrak{R}|} \right\rfloor$. Obviously, there always exists $N_0' \in \mathbb{N}^+$ such that, for any $0 \neq \mathfrak{I} \le_l \mathfrak{R}$ and $\dfrac{\log|\mathfrak{I}|}{\log|\mathfrak{R}|} \dfrac{R - R_0}{2} > \eta > 0$,

$$\min\left\{ H(\mathbf{S}_{\mathfrak{R}/\mathfrak{I}}|\pi), H\left(\mathbf{P}|\pi\right) - \lim_{m\to\infty} \frac{1}{m} H\left(Y^{(m)}_{\mathfrak{R}/\mathfrak{I}}, Y^{(m-1)}_{\mathfrak{R}/\mathfrak{I}}, \cdots, Y^{(1)}_{\mathfrak{R}/\mathfrak{I}} \right) \right\} + \eta - \frac{k}{n} \log|\mathfrak{I}| < -\eta/2 \qquad (14)$$

if $n > N_0'$. 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 $\dfrac{1}{n}\log|\phi(\mathfrak{R}^n)| \le \dfrac{1}{n}\log\left|\mathfrak{R}^k\right| = \dfrac{\log|\mathfrak{R}|}{n}\left\lfloor \dfrac{nR}{\log|\mathfrak{R}|} \right\rfloor \le R.$

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 $\Pr\left\{\psi(\phi(\mathbf{X})) \neq \mathbf{X}\right\} = \Pr\left\{E_1 \cup E_2\right\} < \eta$. First of all, by the AEP of Supremus typicality (Proposition III.2), 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

$$\begin{aligned}
&\Pr\{E_2|\, E_1^c\} \\
&= \sum_{\mathbf{x}'\in\mathcal{S}_\epsilon\setminus\{\mathbf{X}\}} \Pr\{\phi(\mathbf{x}') = \phi(\mathbf{X})|\, E_1^c\} \\
&\le \sum_{0\neq\mathfrak{I}\le_l\mathfrak{R}} \sum_{\mathbf{x}'\in S_\epsilon(\mathbf{X},\mathfrak{I})\setminus\{\mathbf{X}\}} \Pr\{\phi(\mathbf{x}') = \phi(\mathbf{X})|\, E_1^c\} && (15)\\
&< \sum_{0\neq\mathfrak{I}\le_l\mathfrak{R}} \exp_2\left[n(r_{\mathfrak{R}/\mathfrak{I}}+\eta)\right]|\mathfrak{I}|^{-k} && (16)\\
&< \left(2^{|\mathfrak{R}|}-2\right) \max_{0\neq\mathfrak{I}\le_l\mathfrak{R}} \exp_2\left[n\left(r_{\mathfrak{R}/\mathfrak{I}}+\eta-\frac{k}{n}\log|\mathfrak{I}|\right)\right] && (17)\\
&< \left(2^{|\mathfrak{R}|}-2\right)\exp_2(-n\eta/2), && (18)
\end{aligned}$$

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

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

(16) is from the typicality lemmas, Lemma III.3 and Lemma III.4, and [1, Lemma III.3], and it is required that ϵ is smaller than some $\epsilon_0''' > 0$ and n is larger than some $N_0''' \in \mathbb{N}^+$;

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

(18) is from (14), 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 12. From the proof of Theorem IV.1 ([1, Theorem III.1]), one can see that the generalization of the achievability theorem from linear coding technique over finite field to the one over finite ring builds on the generalization of the typicality lemma of Markov sources (the conditional typicality lemma of correlated i.i.d. sources [15, Theorem 15.2.2]) and the analysis of random linear mappings over finite rings [1, Lemma III.3].

The following is an example to help interpreting the above theorems. It is seen from this example that (12), as well as (13), coincides with (1) for $s = 1$.

Example IV.3. 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	.8142	.1773	.0042	.0042
1	.0042	.9873	.0042	.0042
2	.0042	.1773	.8142	.0042
3	.0042	.1773	.0042	.8142

By Theorem IV.1, we have that $\mathcal{R} = \{R \in \mathbb{R} | R > \max\{0.1602, 0.1474\} = H(\mathbf{P}|\pi)\}$, where π is the invariant distribution of $\mathscr{M}$, is achievable with linear coding over $\mathbb{Z}_4$. One can easily see that $\mathcal{R}$ is just the optimal region given by (1) for $s = 1$.

Although the achievable regions presented in the above theorems are comprehensive, they depict the optimal one in many situations, i.e. (13) (or (12)) is identical to (1) for $s = 1$. This has been demonstrated by Example IV.3 above, and more is shown in the following.

Corollary IV.4. *In Theorem IV.1 (or Theorem IV.2), if $\mathfrak{R}$ is a finite field, then*

$$R > H(\mathbf{P}|\pi)$$

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

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 $H\left(Y^{(m)}_{\mathfrak{R}/\mathfrak{R}}\right) = 0$ for all feasible m. ■

Corollary IV.4 says that linear coding over finite fields is always optimal for the special case of Problem 2 considered in this section. However, it is not yet conclusively proved that linear coding over any non-field ring can be equally optimal, other than shown in Example IV.3. Nevertheless, it has been proved that, in the case of multiple i.i.d. correlated sources, there always exist non-field rings over which linear coding is optimal [25]. As a matter of fact, the single source scenario of this assertion is included as a special case of Theorem IV.2 (see Corollary IV.5).

Corollary IV.5. *In Theorem IV.2, 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; or*
2) *$\mathfrak{R}$ 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 $\bigcup_{\Phi \in \mathcal{I}(\mathscr{Y},\mathfrak{R})} \mathcal{R}_\Phi$ *is the Slepian–Wolf region* $\{R \in \mathbb{R}\,|R > H(\pi)\}$.

Proof: The first half of the statement follows from Theorem IV.2 by direct calculation. The second half is from [25]. ■

Remark 13. Concrete examples of the finite ring from Corollary IV.5 includes, but are not limited to:

1) $\mathbb{Z}_p$, where p 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 is a prime;
3) $M_{L,p_1} \times \mathbb{Z}_{p_2}$, where p_1 and p_2 are primes.

Since there always exists a prime p with $p^2 > |\mathscr{Y}|$ in Theorem IV.2, 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..

As mentioned, Corollary IV.5 can be generalized to the multiple sources scenario in a memoryless setting (see [1], [25]). In exact terms, 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 [26]) Markov source(s) scenario holds. Nevertheless, Example IV.3, Corollary IV.5 and [25] do affirmatively support such an assertion to their own extents[7].

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 the problem.

V. Source Coding for Computing Markovian Functions

We are now ready to move on to a more general setting of Problem 2, where both s and g are arbitrary. 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.

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

$$g : (x_1, x_2, x_3) \mapsto x_1 + 2x_2 + 3x_3, \tag{19}$$

and assume that $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$, 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 [12] to first decode the original

[7] The authors conjecture that linear coding claims optimality in the discussed aspect of the problem. However, there may be a weakness in the technique used to obtain (13). This weakness prohibits full extraction of the capability of the linear encoder. Consequently, it could happen that (13) is strictly smaller than (1) for $s = 1$ in some cases.

data and then compute g. This results in an achievable region

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

However, $\mathcal{R}_3$ is not optimal, i.e. coding rates beyond this region can be achieved. Observe that $\left\{Y^{(n)}\right\}_{-\infty}^{\infty}$, 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

(20)

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

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

where $Z_t^k = \phi(X_t^n)$ $(t = 1,2,3)$ 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 also linear. Thus, another approach[8] 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 $\vec{g}(X_1^n, X_2^n, X_3^n)$. Obviously

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

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

$$\mathcal{R}_{\mathbb{Z}_4} = \left\{ [r,r,r] \in \mathbb{R}^3 \middle| r > \frac{2k}{n} = 0.4422 \right\} \tag{21}$$

[8] The idea of this approach is first introduced by Körner and Marton [4] for computing the modulo-two sum of two correlated i.i.d. sources. This is then generalized to the case of arbitrary discrete function based on the observation that any discrete function of finite domain is a restriction of some polynomial function over some finite field [7], [9]. The supports of these approaches are linear coding techniques over finite fields from Elias [27] (binary field) and Csiszár [28] (arbitrary finite field). However, [1] points out that treating an arbitrary discrete function as a polynomial function over some finite ring (instead over field) can lead to strictly better performance. This (encoding polynomial functions over finite rings) requires establishing the achievability theorems, [1, Theorem III.1] and Theorem IV.1, of linear coding techniques over rings.

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" casted 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.

In the rest of this section, we will address (c1) first. The conclusion for (c0) will then follow very naturally after the connection between these two conditions is further detailed.

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 $\left\{Z^{(n)}=\sum_{t\in\mathcal{S}}k_t\left(X_t^{(n)}\right)\right\}_{-\infty}^{\infty}$, respectively. We have*

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

where

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

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

$$\mathcal{R}=\{[R,R,\cdots,R]\in\mathbb{R}^s\,|R>H(\mathbf{P}|\pi)\}\,. \tag{22}$$

Proof: 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 $k>\dfrac{nR_0}{\log|\mathfrak{R}|}$. 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. (22) follows. ■

The following example pictures an explicit settings 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. Thus, [12], which assumes that the ergodic property holds for the sources, 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, 0)	(1, 0, 1)	(1, 1, 0)	(1, 1, 1)
(0, 0, 0)	.1493	.8120	.0193	.0193	0	0	0	0
(0, 0, 1)	.0193	.9420	.0193	.0193	0	0	0	0
(0, 1, 0)	.0193	.8120	.1493	.0193	0	0	0	0
(0, 1, 1)	.0193	.8120	.0193	.1493	0	0	0	0
(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

and

	(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)	0	0	0	0	.1493	.8120	.0193	.0193
(0, 0, 1)	0	0	0	0	.0193	.9420	.0193	.0193
(0, 1, 0)	0	0	0	0	.0193	.8120	.1493	.0193
(0, 1, 1)	0	0	0	0	.0193	.8120	.0193	.1493
(1, 0, 0)	.1493	.8120	.0193	.0193	0	0	0	0
(1, 0, 1)	.0193	.9420	.0193	.0193	0	0	0	0
(1, 1, 0)	.0193	.8120	.1493	.0193	0	0	0	0
(1, 1, 1)	.0193	.8120	.0193	.1493	0	0	0	0

,

respectively. Let $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with $\mathscr{M}$ defined above. It is seen that $\mathscr{M}$ does not process the ergodic property in a strong sense [29, pp. 68], i.e. $\prod_{n=1}^{\infty} \mathbf{P}^{(n)}$ does not tend to a limiting matrix with identical rows. Furthermore, there does also not exist an "invariant distribution" π' such that $\pi'\mathbf{P}^{(n)} = \pi'$ for all feasible n. Therefore, $\mathscr{M}$ is not *asymptotically mean stationary* [30], hence $\mathscr{M}$ possesses no ergodic property [30, Theorem 7.1 and Theorem 8.1]. As a consequence, [12] does not apply. However, it can be easily verified that the function g is still Markovian although $\mathscr{M}$ is not even homogeneous. Moreover, it admits the same stochastic property as shown in Example V.1. In exact terms, $\left\{g\left(X^{(n)}\right)\right\}_{-\infty}^{\infty}$ is homogeneous irreducible Markovian with transition matrix $\mathbf{P}$ given by (20). Consequently, Theorem V.2 offers a solution which achieves (21).

For an arbitrary g, Lemma II.15 promises that there always 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

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

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

Corollary V.4. *In Problem 2, assume that* $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ *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*

$$g(x_1, x_2, \cdots, x_s) = h\left(\sum_{t=1}^{s} k_t(x_t)\right) \tag{23}$$

and $\mathscr{M} = \left\{Z^{(n)} = \sum_{t=1}^{s} k_t\left(X_t^{(n)}\right)\right\}_{-\infty}^{\infty}$ *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*

$$\mathcal{R}_{\mathfrak{R}} = \{[R, R, \cdots, R] \in \mathbb{R}^s | R > R_0\} \subseteq \mathcal{R}[g]. \tag{24}$$

Proof: The existences of k_t's and h are from Lemma II.15, and Lemma C.1 ensures that $\mathscr{M}$ is Markovian. In addition, $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 $\left\{Y_{\mathfrak{R}/\mathfrak{I}}^{(n)}\right\}_{-\infty}^{\infty}$ is Markovian by Lemma C.1. This implies that $\mathcal{R}_{\mathfrak{R}} \subseteq \mathcal{R}[g]$ by Theorem V.2. ■

Remark 14. 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 15. 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

$$\mathcal{R}_s \bigcup \left(\bigcup_{\mathfrak{R}} \bigcup_{\mathscr{P}_{\mathfrak{R}}(g)} \mathcal{R}_{\mathfrak{R}} \right), \tag{25}$$

where $\mathscr{P}_{\mathfrak{R}}(g)$ denotes all the polynomial presentations of format (23) 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, $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ and $\mathscr{M}$ are i.i.d., then*

$$\mathcal{R}_{\mathfrak{R}} = \left\{ [R,R,\cdots,R] \in \mathbb{R}^s \,\middle|\, R > \max_{0\neq\mathfrak{I}\leq_l\mathfrak{R}} \frac{\log|\mathfrak{R}|}{\log|\mathfrak{I}|}\left[H(\pi) - H(\pi_{\mathfrak{I}})\right] \right\} \subseteq \mathcal{R}[g], \tag{26}$$

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

Remark 16. 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], [3], [25].

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 17. 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 always outperforms using a field, from several points of view. 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 achieving $\mathcal{R}[g]$ if the sources are jointly ergodic [12]. 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. [25]. Furthermore, many examples, say Example IV.3, 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], [3], [11]; 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], [11] 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 C.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 handicaps 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

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

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

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

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 C.2 due to Lemma C.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 18. 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 includes $\sum_{t=1}^{s} x_t \in \mathbb{Z}_{2p}[s]$ 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* (25)*, 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 [7], [9], [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. The new techniques are hopefully helpful in understanding and investigating related problems.

APPENDIX A
PROOF OF PROPOSITION II.8

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,

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

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 [16, 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 [16, 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

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

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
Proof of Lemma II.15

Let $\mathbb{F}$ be a finite field such that $|\mathbb{F}|\geq|\mathscr{X}_t|$ for all $1\leq t\leq s$ and $|\mathbb{F}|^s\geq|\mathscr{Y}|$, and let $\mathfrak{R}$ be the *splitting field* of $\mathbb{F}$ of order $|\mathbb{F}|^s$ (one example of the pair $\mathbb{F}$ and $\mathfrak{R}$ is the $\mathbb{Z}_p$, where p is some prime, and its *Galois extension* of *degree* s). It is easily seen that $\mathfrak{R}$ is an s dimensional vector space over $\mathbb{F}$. Hence, there exist s vectors $v_1,v_2,\cdots,v_s\in\mathfrak{R}$ that are linearly independent. Let k_t be an injection from $\mathscr{X}_t$ to the subspace generated by vector v_t. It is easy to verify that $k=\sum_{t=1}^{s}k_t$ is injective since $v_1,v_2,\cdots,v_s$ are linearly independent. Let k' be the inverse mapping of $k:\prod_{t=1}^{s}\mathscr{X}_t\to k\left(\prod_{t=1}^{s}\mathscr{X}_t\right)$ and $\nu:\mathscr{Y}\to\mathfrak{R}$ be any injection. We have that

$$\hat{g}=\nu\circ g\circ k'\in\mathfrak{R}[s]$$

by [22, Lemma 7.40]. Define h to be $\nu'\circ\hat{g}$, where ν' is the inverse mapping of $\nu:\mathscr{Y}\to\nu(\mathscr{Y})$. We have that

$$g=\nu'\circ(\nu\circ g\circ k')\circ k=\nu'\circ\hat{g}\circ k=h\circ k.$$

The statement is proved.

Remark 19. In the proof, k is chosen to be injective because the proof includes the case that g is an identity function. In general, k is not necessarily injective.

Appendix C
Supporting Lemmas

Lemma C.1. *Let $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ 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 $\left\{\Gamma\left(X^{(n)}\right)\right\}_{-\infty}^{\infty}$ 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, $\left\{\Gamma\left(X^{(n)}\right)\right\}_{-\infty}^{\infty}$ is Markov. ■

Remark 20. Lemma C.1 is enlightened by [13, Theorem 3]. However, $\left\{X^{(n)}\right\}_{-\infty}^{\infty}$ in this lemma is not necessary stationary or finite-state.

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

$$\frac{p(z_1, y_2)}{p(z_1)} = c_{y_1, y_2}, \forall\, z_1 \in h^{-1}(y_1), \tag{27}$$

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

$$H\left(h\left(Z^{(2)}\right)\Big|h\left(Z^{(1)}\right)\right) \leq H(\mathbf{P}|\pi), \tag{28}$$

where $\left(Z^{(1)}, Z^{(2)}\right) \sim \pi\mathbf{P}$. Moreover, (28) *holds with equality if and only if*

$$p(z_1, h(z_2)) = p(z_1, z_2), \forall\, z_1, z_2 \in \mathscr{Z} \text{ with } p(z_1, z_2) > 0. \tag{29}$$

Proof: By definition,

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

where (a) is from (27). In addition, equality holds, i.e. (b) holds with equality, if and only if (29) is satisfied.

■

Remark 21. $\mathbf{P}$ in the above lemma can be interpreted as the transition matrix of some Markov process. However, π is not necessary the corresponding invariant distribution. It is also not necessary that such a Markov process is irreducible. In the meantime, (28) can be seen as a "data processing inequality". In addition, (27) is sufficient but not necessary for (28), even though it is sufficient and necessary for (a) in the above proof.

Lemma C.3. *For g given by* (19) *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 $h : \mathscr{S} \to h(\mathscr{S})$. Obviously, h' is bijective. By (19), 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

$$\begin{aligned} &\tau = k_2(0) - k_2(1) = k_2(1) - k_2(0) = -\tau \\ \Longrightarrow\ &\tau + \tau = 0. \end{aligned} \tag{30}$$

(30) 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 D
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 D.1. *Given an irreducible Markov chain $\mathscr{M} = \left\{X^{(n)}\right\}_{-\infty}^{\infty}$ with finite state space $\mathscr{X}$, transition matrix $\mathbf{P}$ and invariant distribution $\pi = [p_j]_{j \in \mathscr{X}}$. 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 $\forall\ \mathbf{x} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right] \in \mathcal{S}_\epsilon(n, \mathbf{P})$,*

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

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 $\mathbf{y} = \left[y^{(1)}, y^{(2)}, \cdots, y^{(n)}\right] \in S_\epsilon(\mathbf{x})$, it is easily seen that

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

is a strongly Markov ϵ-typical sequence of length m_k with respect to $\mathbf{S}_k$, since $\mathbf{y}$ is Supremus ϵ-typical. Additionally, by Proposition II.8, 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 $\exp_2\{m_k\left[H(\mathbf{S}_k|\pi_k) + \eta/2\right]\}$ 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

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

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

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

(8) is established. Direct calculation yields (9). ■

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

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

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

Proof: 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* [12] 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. ■

References

[1] S. Huang and M. Skoglund, "On linear coding over finite rings and applications to computing," *IEEE Transactions on Information Theory*, Submitted. [Online]. Available: http://www.ee.kth.se/~sheng11

[2] 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.

[3] S. Huang and M. Skoglund, "On achievability of linear source coding over finite rings," in *IEEE International Symposium on Information Theory*, July 2013.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

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

[11] ——, *On Existence of Optimal Linear Encoders over Non-field Rings for Data Compression with Application to Computing*, KTH Royal Institute of Technology. [Online]. Available: http://www.ee.kth.se/~sheng11

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

[13] 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, ArticleType: research-article / Full publication date: Dec., 1958 / Copyright © 1958 Institute of Mathematical Statistics. [Online]. Available: http://www.jstor.org/stable/2236949

[14] 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

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

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

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

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

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

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

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

[22] R. Lidl and H. Niederreiter, *Finite Fields*, 2nd ed. New York: Gambridge University Press, 1997.

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

[24] 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.

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

[26] 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, p. 5 pp. vol.3.

[27] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, pp. 37–46, Mar. 1955.

[28] 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.

[29] J. Hajnal and M. S. Bartlett, "The ergodic properties of non-homogeneous finite markov chains," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, no. 01, pp. 67–77, 1956.

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