Finite-State Mutual Dimension*

Adam Case[1] and Jack H. Lutz[2]

***Abstract*—In 2004, Dai, Lathrop, Lutz, and Mayordomo defined and investigated the *finite-state dimension* (a finite-state version of algorithmic dimension) of a sequence S and, in 2018, Case and Lutz defined and investigated the *mutual (algorithmic) dimension* between two sequences S and T. In this paper, we propose a definition for the *lower* and *upper finite-state mutual dimensions* $mdim_{FS}(S:T)$ and $Mdim_{FS}(S:T)$ between two sequences S and T. Intuitively, the finite-state dimension of a sequence S represents the density of finite-state information contained within S, while the finite-state mutual dimension between two sequences S and T represents the density of finite-state information *shared* by S and T. Thus "finite-state mutual dimension" can be viewed as a "finite-state" version of mutual dimension and as a "mutual" version of finite-state dimension.**

The main results of this investigation are as follows. First, we show that finite-state mutual dimension, defined using *information-lossless finite-state compressors*, has all of the properties expected of a measure of *mutual information*. Next, we prove that finite-state mutual dimension may be characterized in terms of *block mutual information rates*. Finally, we provide necessary and sufficient conditions for two *normal* sequences R_1 and R_2 to achieve $mdim_{FS}(R_1 : R_2) = Mdim_{FS}(R_1 : R_2) = 0$.

***Index Terms*—finite-state compression, Shannon entropy, mutual information, finite-state dimension, normality, independence**

I. Introduction

The study of *algorithmic dimension* has yielded various mechanisms for quantifying the density of information contained within infinite objects, such as points in Euclidean space [19] and sequences [15]. Recent investigations into the dimensions of points and sequences have produced new characterizations of classical Hausdorff dimension [11, 17, 18] and insights into self-similar fractal geometry [8, 10, 19], among other results. Originally defined in terms of *gales* (a generalization of *martingales*) [15], the *dimension* $dim(S)$ and *strong dimension* $Dim(S)$ of a sequence $S \in \Sigma^\infty$ were shown to have the characterizations

$$dim(S) = \liminf_{n\to\infty} \frac{K(S \upharpoonright n)}{n \log |\Sigma|}$$

and

$$Dim(S) = \limsup_{n\to\infty} \frac{K(S \upharpoonright n)}{n \log |\Sigma|},$$

where $K(S \upharpoonright n)$ is the *Kolmogorov complexity* of the first n symbols of S [2, 20]. These characterizations show that $dim(S)$ and $Dim(S)$ can be thought of as the lower and upper densities of *algorithmic information* contained within S. The algorithmic dimension and *algorithmic randomness* of sequences have been shown to have interesting relationships. For example, if a sequence $S \in \Sigma^\infty$ is *(algorithmically) random*, then $dim(S) = 1$. However, not all sequences that achieve $dim(S) = 1$ are necessarily random [15].

The notion of the dimension of a sequence has been adapted to operate within different contexts in the fields of computability and information theory. For example, Dai, Lathrop, Lutz, and Mayordomo developed the notion of *finite-state dimension*, which is a finite-state version of algorithmic dimension [7]. In their paper, the authors define finite-state dimension in terms of *finite-state gamblers*. In [7] and [2] the authors show that the *finite-state dimension* $dim_{FS}(S)$ and *finite-state strong dimension* $Dim_{FS}(S)$ of a sequence $S \in \Sigma^\infty$ may be characterized by

$$dim_{FS}(S) = \lim_{r\to\infty} \liminf_{n\to\infty} \rho_r(S \upharpoonright n)$$

and (1)

$$Dim_{FS}(S) = \lim_{r\to\infty} \limsup_{n\to\infty} \rho_r(S \upharpoonright n),$$

where $\rho_r(S \upharpoonright n)$ is defined by

$$\rho_r(S \upharpoonright n) = \min\left\{ \frac{|C(S \upharpoonright n)|}{n \log |\Sigma|} \,\middle|\, C \text{ is an ILFSC that has } r \text{ states} \right\},$$

and C is an *information-lossless finite-state compressor* (ILFSC) and $|C(S \upharpoonright n)|$ is the length of the output that C produces when given the first n symbols of S as input. These quantities can be thought of as the lower and upper densities of *finite-state information* contained within S and are also known as the *lower* and *upper compression ratios* of S as studied by Ziv and Lempel [23].

Other characterizations of finite-state dimension have been shown. For example, Bourke, Hitchcock, and Vinodchandran proved that the lower and upper finite-state dimensions of a sequence $S \in \Sigma^\infty$ are equal to the lower and upper *block entropy rates* of S, respectively (i.e., the lower and upper limiting normalized entropies of the frequencies of aligned blocks of symbols contained within S) [4]. In a recent paper, Kozachinskiy and Shen show that finite-state dimension can also be characterized in terms of the entropy rates of non-aligned blocks of symbols and in terms of superadditive calibrated functions on strings [13].

There have been several interesting explorations into the relationships between finite-state dimension and the concept of *normality*, which was introduced by Borel in 1909 [9]. A sequence $S \in \Sigma^\infty$ is *normal* if every string of the same length occurs with the same limiting frequency within S. Normality

*The second author's research was supported in part by National Science Foundation grants 1545028 and 1900716. Both authors thank the National University of Singapore for their support during the program Equidistribution: Arithmetic, Computational and Probabilistic Aspects at the Institute for Mathematical Sciences, where part of this research was carried out.

[1]Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311, USA adam.case@drake.edu

[2]Department of Computer Science, Iowa State University, Ames, IA 50011, USA lutz@iastate.edu

can be viewed as a weaker form of randomness, since every algorithmically random sequence is also normal but not vice-versa. In fact, it has been shown that a sequence $S \in \Sigma^\infty$ is normal if and only if $dim_{FS}(S) = 1$ [4, 7]. Thus the normal sequences can be completely characterized as the sequences that achieve finite-state dimension one. This equivalence has recently been quantitatively refined using the Kullback-Leibler divergence [12].

Another way in which the dimensions of sequences have been adapted to fit other contexts within information theory can be found in the development of *mutual dimension*, which was introduced in 2015 by the present authors in [5]. In this paper, the authors defined the mutual dimension between two points in Euclidean space and showed that it has all the properties expected of a measure of mutual information, including several *data processing inequalities*. In 2018, the same authors extended this framework to sequences and defined the *lower* and *upper mutual dimensions*, $mdim(S : T)$ and $Mdim(S : T)$, respectively, between two sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ by

$$mdim(S) = \liminf_{n\to\infty} \frac{I(S \upharpoonright n : T \upharpoonright n)}{n \log |\Sigma|}$$

and

$$Mdim(S) = \limsup_{n\to\infty} \frac{I(S \upharpoonright n : T \upharpoonright n)}{n \log |\Sigma|},$$

where $I(S \upharpoonright n : T \upharpoonright n)$ is the *algorithmic mutual information* between the first n bits of S and T [6]. The algorithmic mutual information $I(u : w)$ between two strings $u \in \Sigma^*$ and $w \in \Sigma^*$ is

$$I(u : w) = K(w) - K(w \mid u),$$

where $K(w \mid u)$ is the *Kolmogorov complexity* of w *given* u. However, this quantity can also be characterized by

$$I(u : w) = K(u) + K(w) - K(u, w) + o(|u|), \qquad (2)$$

where $K(u, w)$ is the *joint Kolmogorov complexity* of u and w. (The interested reader may refer to [14] for an in-depth discussion on algorithmic mutual information.) Therefore, we can view the lower and upper mutual dimensions as the lower and upper densities of algorithmic information *shared* by two sequences. In the same paper, the authors demonstrate that, if two sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are *independently random*, then $Mdim(S : T) = 0$. However, they also show that not all pairs of sequences that achieve mutual dimension zero are necessarily independently random [6].

The goal of this article is to develop a notion of *finite-state mutual dimension* (which includes using information-lossless finite-state compressors in the definition) in order to provide a mechanism in which to reason about the "mutual compressibility" of two sequences. Given the quantities found in (1) and (2), a natural way one might define the *lower* and *upper finite-state mutual dimensions* between sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ is

$$\lim_{r\to\infty} \liminf_{n\to\infty} \left[\rho_r(S \upharpoonright n) + \rho_r(T \upharpoonright n) - \rho_r(S \upharpoonright n, T \upharpoonright n)\right]$$

and

$$\lim_{r\to\infty} \limsup_{n\to\infty} \left[\rho_r(S \upharpoonright n) + \rho_r(T \upharpoonright n) - \rho_r(S \upharpoonright n, T \upharpoonright n)\right],$$

respectively. However, it is unclear whether or not the limits in the proposed definitions above even exist and, if they do, whether or not these definitions possess the appropriate properties for a measure of mutual information. In this article, we take a different (and less natural) approach to developing these definitions, one that makes use of *iterated limits* as the number of states goes to infinity. Ultimately, we will demonstrate the robustness of these definitions by relating them to *block entropy rates* and to the concept of *normality*.

The outline of this article is as follows. In Section II, we discuss the *Shannon entropy* of a particular class of probability measures that quantify the *block frequencies* of strings. Using Ziv and Lempel's Generalized Kraft Inequality [23], we are able to establish upper bounds on the *difference* between the normalized entropy of these probability measures and the *compression ratio* of strings. We use these bounds to prove the basic properties of the *mutual compression ratio* between two strings. In Section III, we extend the notion of the mutual compression ratio to infinite sequences and use it to define the lower and upper finite-state mutual dimensions. We prove an important theorem regarding the interchangeability of the iterated limits within the definition of finite-state mutual dimension, which we then use to prove the properties of finite-state mutual dimension. In Section IV, we introduce the *lower* and *upper block mutual information rates* between two sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ and show that they are equal to the lower and upper finite-state mutual dimensions, respectively. In Section V, we obtain a result regarding the *independence* of sequences at the finite-state level. Specifically, we prove that, if $R_1 \in \Sigma^\infty$ and $R_2 \in \Sigma^\infty$ are normal, then the sequence $(R_1, R_2) \in (\Sigma \times \Sigma)^\infty$ is normal if and only if $Mdim_{FS}(R_1 : R_2) = 0$, where (R_1, R_2) is the sequence obtained by pairing the symbols of R_1 and R_2 at the same index.

II. Entropy, Block Frequencies, and Mutual Compression

In this section, we define and investigate the *mutual compression ratio* between two strings. To do this, we make use of some relationships between compression ratios and entropies of the relative frequencies of strings that were originally established by Ziv and Lempel [23] and further examined by Sheinwald [22].

In this paper, we assume that Σ is an alphabet consisting of k symbols. We write Σ^* to represent the set of all strings over Σ and Σ^∞ to represent the set of all sequences over Σ. The *length* of a string $u \in \Sigma^*$ is denoted by $|u|$ and we represent the set of all strings of length $n \in \mathbb{N}$ by Σ^n. The *empty string* (the string of length zero) is denoted by λ. For any sequence $S \in \Sigma^\infty$, we write $S \upharpoonright n$ for the first $n \in \mathbb{N}$ symbols of S.

For any string $u \in \Sigma^*$ and sequence $S \in \Sigma^\infty$, we write $u[i]$ and $S[i]$ for the i^{th} symbol of u and the i^{th} symbol of S, respectively. For any two strings $u \in \Sigma^n$ and $w \in \Sigma^n$, we write (u, w) to represent the string

$$(u, w) = (u[1], w[1])(u[2], w[2]) \cdots (u[n], w[n]) \in (\Sigma \times \Sigma)^n.$$

Note that the lengths of u and w must be equal in order to use the notation (u, w) for strings. Similarly, for any two sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$, we write (S, T) to represent the sequence

$$(S, T) = (S[1], T[1])(S[2], T[2]) \cdots \in (\Sigma \times \Sigma)^\infty.$$

We will write $\log$ for the base-2 logarithm function and $\log_k$ for the base-k logarithm function.

A *discrete probability measure* α on a finite set $\mathcal{X}$ is a function $\alpha : \mathcal{X} \to [0, 1]$ such that

$$\sum_{x \in \mathcal{X}} \alpha(x) = 1.$$

Definition. Let α be a discrete probability measure on $\mathcal{X}$. The *Shannon entropy* of α is

$$\mathrm{H}(\alpha) = \sum_{x \in \mathcal{X}} \alpha(x) \log \frac{1}{\alpha(x)}.$$

If α is a discrete probability measure on $\mathcal{X} \times \mathcal{X}$ (sometimes called a *joint probability measure* on $\mathcal{X}$), we will write $\alpha(x, y)$ to denote the value $\alpha((x, y))$ assigned to the pair (x, y) by α. The *first* and *second marginal probability measures* of α are the probability measures α_1 and α_2 on $\mathcal{X}$ defined by

$$\alpha_1(a) = \sum_{b \in \mathcal{X}} \alpha(a, b) \quad \text{and} \quad \alpha_2(b) = \sum_{a \in \mathcal{X}} \alpha(a, b),$$

respectively.

For any $n, \ell \in \mathbb{Z}^+$ such that n is a multiple of ℓ and all $x \in \Sigma^\ell$ and $u \in \Sigma^n$, we denote the number of *block occurrences* of x in u by

$$\#_\Box(x, u) = \left| \left\{ m \leq \frac{|u|}{|x|} \,\middle|\, u[m|x| + 1 \ldots (m + 1)|x|] = x \right\} \right|,$$

where $u[i \ldots j]$ is the substring of u starting at index i and ending at index j, for all $i, j \in \mathbb{Z}^+$ such that $i \leq j$. We represent the *block frequency* of x in u by $\pi_u(x)$, where the function $\pi_u : \Sigma^* \to \mathbb{Q}_{[0,1]}$ is defined by

$$\pi_u(x) = \frac{l}{n} \#_\Box(x, u)$$

and $\mathbb{Q}_{[0,1]}$ is the set of all rationals in $[0, 1]$. For all $n, \ell \in \mathbb{Z}^+$ such that n is a multiple of ℓ and $u \in \Sigma^n$, we denote the restriction of π_u to strings in Σ^ℓ by $\pi_u^{(\ell)}$. It is important to note that $\pi_u^{(\ell)}$ represents a discrete probability measure on the finite set Σ^ℓ.

For all $x, y \in \Sigma^\ell$ and $u, w \in \Sigma^n$, we represent the *joint block frequency* of x in u and y in w by $\pi_{u,w}(x, y)$, where the function $\pi_{u,w} : \Sigma^* \times \Sigma^* \to \mathbb{Q}_{[0,1]}$ is defined by

$$\pi_{u,w}(x, y) = \frac{\ell}{n} \#_\Box((x, y), (u, w)).$$

We denote the restriction of $\pi_{u,w}$ to the pairs of strings in $\Sigma^\ell \times \Sigma^\ell$ by $\pi_{u,w}^{(\ell)}$. Once again, we note that $\pi_{u,w}^{(\ell)}$ is a discrete probability measure on $\Sigma^\ell \times \Sigma^\ell$. It is easy see that, for all $x, y \in \Sigma^\ell$, $\pi_{(u,w)}^{(\ell)}((x, y)) = \pi_{u,w}^{(\ell)}(x, y)$. Also, it is important to observe that the first and second marginal probability measures of $\pi_{u,w}^{(\ell)}$ are $\pi_u^{(\ell)}$ and $\pi_w^{(\ell)}$, respectively.

To prove the main theorem of this section, we must examine relationships between the Shannon entropy $\mathrm{H}(\pi_u^\ell)$ of the probability measure π_u^ℓ and the finite-state *compressibility* of $u \in \Sigma^*$. We will also revisit the entropy of block frequencies and joint block frequencies of strings in Section IV.

A *finite-state compressor* (FSC) C on Σ is a 4-tuple

$$C = (Q, \delta, \nu, q_0),$$

where Q is a nonempty finite set of *states*, $\delta : Q \times \Sigma \to Q$ is the *transition function*, $\nu : Q \times \Sigma \to \{0, 1\}^*$ is the *output function*, and q_0 is the *initial state*. We define the *extended transition function* $\delta^* : Q \times \Sigma^* \to Q$ by the recursion

$$\delta^*(q, \lambda) = q,$$
$$\delta^*(q, ua) = \delta(\delta^*(q, u), a),$$

for all $q \in Q$, $u \in \Sigma^*$, and $a \in \Sigma$. The output function ν is defined by the recursion

$$\nu(q, \lambda) = \lambda,$$
$$\nu(q, ua) = \nu(q, u)\nu(\delta^*(q, u), a),$$

for all $q \in Q$, $u \in \Sigma^*$ and $a \in \Sigma$. The *output* of C on the input string $u \in \Sigma^*$ is denoted by $C(u) = \nu(q_0, u)$. An *information-lossless finite-state compressor* (ILFSC) C is an FSC where the function $f : \Sigma^* \to \{0, 1\}^* \times Q$, defined by $f(u) = (C(u), \delta^*(q_0, u))$, is one-to-one. Thus we are able to recover the input string u given knowledge of the output string $C(u)$ and the last state C entered $\delta^*(q_0, u)$ after processing u from the initial state q_0. (We note that there exist other equivalent notions of information-losslessness. For example, Ziv and Lempel used a slightly different definition that allows the FSC to start in any state, and when given knowledge of the output string, the last state entered, and the starting state, the input string can be recovered. Our definition has been used by other authors (e.g., [4, 7]) and is equivalent to the one provided by Ziv and Lempel. Proving this equivalence is a simple exercise that we leave to the reader.)

The *compression ratio of* $u \in \Sigma^n$ attained by an ILFSC C on Σ is

$$\rho_C(u) = \frac{|C(u)|}{n \log k}.$$

Likewise, the *joint compression ratio of* $u \in \Sigma^n$ and $w \in \Sigma^n$ attained by an ILFSC C on $\Sigma \times \Sigma$ is

$$\rho_C(u, w) = \frac{|C((u, w))|}{n \log k}.$$

Definition. The *r-state compression ratio of* $u \in \Sigma^n$ is

$$\rho_r(u) = \min \{\rho_C(u) \,|\, C \text{ is an ILFSC on } \Sigma \text{ that has } r \text{ states}\}.$$

Definition. The *r-state joint compression ratio of* $u \in \Sigma^n$

and $w \in \Sigma^n$ is

$$\rho_r(u, w) =$$

$$\min\left\{\rho_C(u, w) \,|\, C \text{ is an ILFSC on } \Sigma \times \Sigma \text{ that has } r \text{ states}\right\}.$$

It is important to note that $\rho_r((u, w))$ is the r-state compression ratio of the string $(u, w) \in (\Sigma \times \Sigma)^n$ and $\rho_r(u, w)$ is the r-state joint compression ratio of $u \in \Sigma^n$ and $w \in \Sigma^n$. Observe that $\rho_r(u, w) = 2\rho_r((u, w))$, since the definition of $\rho_r(u, w)$ divides the compressor's output by $n \log k$ and the definition of $\rho_r((u, w))$ divides the compressor's output by $n \log k^2$.

We proceed to define the *mutual compression ratio* between two strings and explore its properties.

Definition. Let $r, t \in \mathbb{Z}^+$. The *r, t-state mutual compression ratio* between $u \in \Sigma^n$ and $w \in \Sigma^n$ is

$$\rho_{r,t}(u : w) = \rho_t(u) + \rho_t(w) - \rho_r(u, w).$$

The main theorem of this section (Theorem 4) lists the properties of r, t-state mutual compression ratios between strings. Proving this theorem requires that we examine relationships between the individual compression ratios of the strings $u \in \Sigma^*$ and $w \in \Sigma^*$ and the joint compression ratio of u and w. We develop these relationships by converting compression ratios to entropies of block frequencies (and vice-versa) while taking into account negligible error terms. To do this, we make use of an important lemma from Ziv and Lempel.

Lemma 1 (Generalized Kraft Inequality [23]). *For any ILFSC C on Σ with a state set $Q = \{q_1, q_2, \ldots, q_s\}$,*

$$\sum_{w \in \Sigma^r} 2^{-L_C(w)} \leq s^2\left(1 + \log \frac{s^2 + k^r}{s^2}\right),$$

where

$$L_C(w) = \min_{q \in Q}\{|C_q(w)|\}$$

and C_q is the ILFSC that is like C except that it uses q as the initial state.

We also make use of the following inequality that was noted by Sheinwald in [22]. Originally, Ziv and Lempel noted a similar inequality in [23].

Lemma 2 (Sheinwald [22]). *Let C be an ILFSC on Σ. For every $\ell, n \in \mathbb{Z}^+$ and $u \in \Sigma^n$ such that n is a multiple of ℓ,*

$$\rho_C(u) \geq \frac{1}{\ell \log k} \sum_{x \in \Sigma^\ell} \pi_u^{(\ell)}(x) L_C(x).$$

It is worth noting that Ziv and Lempel and Sheinwald originally used the notation $P(x, u)$ in place of $\pi_u^{(\ell)}(x)$.

Using Lemmas 1 and 2, we are able to establish the following upper bound on the *difference* between the normalized entropy of the block frequencies of strings and the compression ratio of strings.

Lemma 3. *Let C be an ILFSC on Σ with $s \in \mathbb{Z}^+$ states. For every $\ell, n \in \mathbb{N}$ and $u \in \Sigma^n$ such that $\ell \leq n$,*

$$\frac{\mathrm{H}(\pi_{u_\ell}^{(\ell)})}{\ell \log k} - \rho_C(u) \leq \left\lfloor \frac{n}{\ell} \right\rfloor^{-1} + f_s^k(\ell),$$

where $u_\ell = u \upharpoonright \left\lfloor \frac{n}{\ell} \right\rfloor \cdot \ell$ and $\lim_{m\to\infty} f_s^k(m) = 0$.

Using this upper bound (and other similar bounds) we are able to prove the main theorem of this section.

Theorem 4 (Properties of Mutual Compression Ratios). *For every $r, t, n \in \mathbb{Z}^+$ and every $u \in \Sigma^n$ and $w \in \Sigma^n$ such that $n \geq \max\{t', r'\}$,*

1) $\rho_{r,t}(u : w) \leq \min\{\rho_t(u), \rho_t(w)\} + \left\lfloor \frac{n}{t'} \right\rfloor^{-1} + g_r^k(t')$ *and* $\lim\limits_{m\to\infty} g_r^k(m) = 0$,
2) $\rho_{r,t}(u : w) + 2\left\lfloor \frac{n}{r'} \right\rfloor^{-1} + h_t^k(r') \geq 0$ *and* $\lim\limits_{m\to\infty} h_t^k(m) = 0$,
3) $\rho_{r,t}(u : u) + \left\lfloor \frac{n}{r'} \right\rfloor^{-1} + i_t^k(r') \geq \rho_t(u)$ *and* $\lim\limits_{m\to\infty} i_t^k(m) = 0$,
4) $\rho_{r,t}(u : u) \leq \rho_t(u) + \left\lfloor \frac{n}{t'} \right\rfloor^{-1} + j_r^k(t')$ *and* $\lim\limits_{m\to\infty} j_r^k(m) = 0$,
5) $\rho_{r,t}(u : w) = \rho_{r,t}(w : u)$, *and*
6) $\rho_{r,t}(u : w) \leq \rho_{t,r}(u : w) + 3\left\lfloor \frac{n}{t'} \right\rfloor^{-1} + e_r^k(t')$ *and* $\lim\limits_{m\to\infty} e_r^k(m) = 0$,

where $r' = \lfloor \log_k r \rfloor$ and $t' = \lfloor \log_k t \rfloor$.

III. Finite-State Mutual Dimension

In this section we define the *lower* and *upper mutual compression ratios* and the *lower* and *upper finite-state mutual dimensions* between sequences and explore their properties.

We begin by discussing the *finite-state dimension* $dim_{FS}(S)$ of a sequence $S \in \Sigma^\infty$, which was originally defined in 2003 by Dai, Lathrop, Lutz, and Mayordomo in [7] using finite-state gamblers. In the same paper, the authors proved a characterization of finite-state dimension using finite-state compressors. In 2007, Athreya, Hitchcock, Lutz, and Mayordomo defined the *finite-state strong dimension* $Dim_{FS}(S)$ of a sequence $S \in \Sigma^\infty$ using finite-state gamblers and proved that it can also be characterized using finite-state compressors [2]. In this section, we will use the compressor characterization of finite-state dimension and finite-state strong dimension and refer to them as the *lower* and *upper finite-state dimensions*, respectively.

We proceed to discuss compression ratio characterizations of the lower and upper finite-state dimensions.

Definition. Let $r \in \mathbb{Z}^+$. The *lower* and *upper r-state compression ratios* of $S \in \Sigma^\infty$ are

$$\rho_r(S) = \liminf_{n\to\infty} \rho_r(S \upharpoonright n)$$

and

$$\hat{\rho}_r(S) = \limsup_{n\to\infty} \rho_r(S \upharpoonright n),$$

respectively.

Definition. The *lower* and *upper finite-state compression ratios* of $S \in \Sigma^\infty$ are

$$\rho(S) = \lim_{r\to\infty} \rho_r(S)$$

and

$$\hat{\rho}(S) = \lim_{r\to\infty} \hat{\rho}_r(S),$$

respectively.

In the following theorem, the first equality was proven by Dai, Lathrop, Lutz, and Mayordomo in [7] and the second equality was proven by Athreya, Hitchcock, Lutz, and Mayordomo in [2].

Theorem 5 ([2, 7])**.** *For all* $S,T \in \Sigma^\infty$*,*

$$dim_{FS}(S) = \rho(S)$$

and

$$Dim_{FS}(S) = \hat{\rho}(S).$$

We now define "joint" versions of finite-state compression ratios.

Definition. Let $r \in \mathbb{Z}^+$. The *lower* and *upper* r*-state joint compression ratios* of $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\rho_r(S,T) = \liminf_{n\to\infty} \rho_r(S \upharpoonright n, T \upharpoonright n)$$

and

$$\hat{\rho}_r(S,T) = \limsup_{n\to\infty} \rho_r(S \upharpoonright n, T \upharpoonright n),$$

respectively.

Definition. The *lower* and *upper joint finite-state compression ratios* of $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\rho(S,T) = \lim_{r\to\infty} \rho_r(S,T)$$

and

$$\hat{\rho}(S,T) = \lim_{r\to\infty} \hat{\rho}_r(S,T),$$

respectively.

The following corollary follows directly from Theorem 5.

Corollary 6. *For all* $S,T \in \Sigma^\infty$*,*

$$dim_{FS}(S,T) = \rho(S,T)$$

and

$$Dim_{FS}(S,T) = \hat{\rho}(S,T).$$

Next, we introduce the *lower* and *upper* r,t*-state mutual compression ratios* between sequences.

Definition. Let $r,t \in \mathbb{Z}^+$. The *lower* and *upper* r,t*-state mutual compression ratios* between $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\rho_{r,t}(S:T) = \liminf_{n\to\infty} \rho_{r,t}(S \upharpoonright n : T \upharpoonright n)$$

and

$$\hat{\rho}_{r,t}(S:T) = \limsup_{n\to\infty} \rho_{r,t}(S \upharpoonright n : T \upharpoonright n),$$

respectively.

We now present the definitions of the *lower* and *upper finite-state mutual dimensions* between sequences.

Definition. The *lower* and *upper finite-state mutual dimensions* between $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$mdim_{FS}(S:T) = \lim_{r\to\infty}\lim_{t\to\infty} \rho_{r,t}(S:T)$$

and

$$Mdim_{FS}(S:T) = \lim_{r\to\infty}\lim_{t\to\infty} \hat{\rho}_{r,t}(S:T),$$

respectively.

The first limit in the definitions above exists because both $\rho_{r,t}(S:T)$ and $\hat{\rho}_{r,t}(S:T)$ are decreasing in t since $\rho_t(S \upharpoonright n)$ and $\rho_t(T \upharpoonright n)$ are decreasing in t. The second limit also exists because both

$$\lim_{t\to\infty} \rho_{r,t}(S:T) \text{ and } \lim_{t\to\infty} \hat{\rho}_{r,t}(S:T)$$

are increasing in r, since $-\rho_r(S \upharpoonright n, T \upharpoonright n)$ is increasing in r.

Our first theorem of this section is an important result that allows for the interchanging of the iterated limits within the definitions of the lower and upper finite-state mutual dimensions. The proof of the properties of finite-state mutual dimensions (Theorem 8) rely on this result.

Theorem 7. *For all* $S,T \in \Sigma^\infty$*,*

$$\begin{aligned} mdim_{FS}(S:T) &= \lim_{r\to\infty}\lim_{t\to\infty} \rho_{r,t}(S:T) \\ &= \lim_{t\to\infty}\lim_{r\to\infty} \rho_{r,t}(S:T) \end{aligned}$$

and

$$\begin{aligned} Mdim_{FS}(S:T) &= \lim_{r\to\infty}\lim_{t\to\infty} \hat{\rho}_{r,t}(S:T) \\ &= \lim_{t\to\infty}\lim_{r\to\infty} \hat{\rho}_{r,t}(S:T). \end{aligned}$$

The final theorem of this section describes the basic properties of finite-state mutual dimension.

Theorem 8 (Properties of Finite-State Mutual Dimensions)**.** *For all* $S,T \in \Sigma^\infty$*,*

1) $mdim_{FS}(S:T) \geq dim_{FS}(S) + dim_{FS}(T) - Dim_{FS}(S,T)$, $mdim_{FS}(S:T) \leq Dim_{FS}(S) + Dim_{FS}(T) - Dim_{FS}(S,T)$.
2) $MDim_{FS}(S:T) \geq dim_{FS}(S) + dim_{FS}(T) - dim_{FS}(S,T)$, $MDim_{FS}(S:T) \leq Dim_{FS}(S) + Dim_{FS}(T) - dim_{FS}(S,T)$.
3) $mdim_{FS}(S:T) \leq \min\{dim_{FS}(S), dim_{FS}(T)\}$, $Mdim_{FS}(S:T) \leq \min\{Dim_{FS}(S), Dim_{FS}(T)\}$.
4) $0 \leq mdim_{FS}(S:T) \leq Mdim_{FS}(S:T) \leq 1$.
5) $mdim_{FS}(S:S) = dim_{FS}(S)$, $Mdim_{FS}(S:S) = Dim_{FS}(S)$.
6) $mdim_{FS}(S:T) = mdim_{FS}(T:S)$, $Mdim_{FS}(S:T) = Mdim_{FS}(T:S)$.

IV. BLOCK MUTUAL INFORMATION RATES

In this section, we introduce the notion of *block mutual information rates* between sequences and prove that the lower and upper finite-state mutual dimensions can be characterized in terms of block mutual information rates.

Originally, Ziv and Lempel proved that the upper finite-state compression ratio of a sequence may be characterized in terms of the *entropy rates* of *non-aligned* block frequencies [23] within the sequence. Sheinwald proved a similar characterization of the upper compression ratio using the *entropy rates* of *aligned* block frequencies [22]. Later, Bourke, Hitchcock, and Vindochandran proved a characterization of the lower and upper finite-state dimensions of sequences [4] in terms of (aligned) block entropy rates. Kozachinskiy and Shen recently proved that the lower finite-state dimension can also be characterized using the entropy rates of non-aligned block frequencies [13].

We begin by discussing the block entropy rates of sequences. For any $n, m \in \mathbb{Z}^+$, $x \in \Sigma^m$, and $S \in \Sigma^\infty$, we denote the n^{th} *block frequency* of x in S by the function $\pi_{S,n} : \Sigma^* \to \mathbb{Q}_{[0,1]}$, defined by

$$\pi_{S,n}(x) = \pi_{S \upharpoonright nm}(x) = \frac{\#_\Box(x, S \upharpoonright nm)}{n}.$$

For each $\ell \in \mathbb{Z}^+$, we denote the restriction of $\pi_{S,n}$ to the strings in Σ^ℓ by $\pi^{(\ell)}_{S,n}$.

Definition. Let $\ell \in \mathbb{Z}^+$. The ℓ^{th} *lower* and *upper block entropy rates of* $S \in \Sigma^\infty$ are

$$\mathrm{H}_\ell(S) = \frac{1}{\ell \log k} \liminf_{n\to\infty} \mathrm{H}(\pi^{(\ell)}_{S,n})$$

and

$$\hat{\mathrm{H}}_\ell(S) = \frac{1}{\ell \log k} \limsup_{n\to\infty} \mathrm{H}(\pi^{(\ell)}_{S,n})$$

respectively.

Definition. The *lower* and *upper block entropy rates of* $S \in \Sigma^\infty$ are

$$\mathrm{H}(S) = \lim_{\ell\to\infty} \mathrm{H}_\ell(S)$$

and

$$\hat{\mathrm{H}}(S) = \lim_{\ell\to\infty} \hat{\mathrm{H}}_\ell(S),$$

respectively.

Using the frameworks developed in [23] and [7], Bourke, Hitchcock, and Vinodchandran proved the following theorem in [4].

Theorem 9 ([4]). *For every* $S \in \Sigma^\infty$,

$$dim_{FS}(S) = \mathrm{H}(S)$$

and

$$Dim_{FS}(S) = \hat{\mathrm{H}}(S).$$

We now define "joint" versions of block entropy rates of sequences. For any $n, m \in \mathbb{Z}^+$, $x, y \in \Sigma^m$, and $S, T \in \Sigma^\infty$, we denote the n^{th} *joint block frequency* of x in S and y in T by the function $\pi_{S,T,n} : \Sigma^* \times \Sigma^* \to \mathbb{Q}_{[0,1]}$, defined by

$$\pi_{S,T,n}(x, y) = \pi_{S\upharpoonright nm, T\upharpoonright nm}(x) = \frac{\#_\Box((x, y), (S, T) \upharpoonright nm)}{n}.$$

As before, for each $\ell \in \mathbb{Z}^+$, we denote the restriction of $\pi_{S,T,n}$ to the pairs of strings in $\Sigma^\ell \times \Sigma^\ell$ by $\pi^{(\ell)}_{S,T,n}$.

Definition. Let $\ell \in \mathbb{Z}^+$. The ℓ^{th} *lower* and *upper joint block entropy rates of* $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\mathrm{H}_\ell(S, T) = \frac{1}{\ell \log k} \liminf_{n\to\infty} \mathrm{H}(\pi^{(\ell)}_{S,T,n})$$

and

$$\hat{\mathrm{H}}_\ell(S, T) = \frac{1}{\ell \log k} \limsup_{n\to\infty} \mathrm{H}(\pi^{(\ell)}_{S,T,n})$$

respectively.

We make note that the ℓ^{th} lower and upper block entropy rates $\mathrm{H}_\ell((S, T))$ and $\hat{\mathrm{H}}_\ell((S, T))$ of $(S, T) \in (\Sigma \times \Sigma)^\infty$ are normalized by $\ell \log k^2$ and the ℓ^{th} lower and upper joint block entropy rates $\mathrm{H}_\ell(S, T)$ and $\hat{\mathrm{H}}_\ell(S, T)$ of $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are normalized by $\ell \log k$.

Definition. The *lower* and *upper joint block entropy rates of* $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\mathrm{H}(S, T) = \lim_{\ell\to\infty} \mathrm{H}_\ell(S, T)$$

and

$$\hat{\mathrm{H}}(S, T) = \lim_{\ell\to\infty} \hat{\mathrm{H}}_\ell(S, T),$$

respectively.

The following corollary follows directly from Theorem 9.

Corollary 10. *For every* $S, T \in \Sigma^\infty$,

$$dim_{FS}(S, T) = \mathrm{H}(S, T)$$

and

$$Dim_{FS}(S, T) = \hat{\mathrm{H}}(S, T).$$

We proceed to introduce the *lower* and *upper block mutual information rates* between sequences. To do this, we make use of Shannon *mutual information*.

Definition. Let α be a discrete probability measure on $\mathcal{X} \times \mathcal{X}$. The *Shannon mutual information* between α_1 and α_2 is

$$\mathrm{I}(\alpha_1; \alpha_2) = \mathrm{H}(\alpha_1) + \mathrm{H}(\alpha_2) - \mathrm{H}(\alpha).$$

Definition. Let $\ell \in \mathbb{Z}^+$. The ℓ^{th} *lower* and *upper block mutual information rates between* $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\mathrm{I}_\ell(S; T) = \frac{1}{\ell \log k} \liminf_{n\to\infty} \mathrm{I}(\pi^{(\ell)}_{S,n}; \pi^{(\ell)}_{T,n})$$

and

$$\hat{\mathrm{I}}_\ell(S;T) = \frac{1}{\ell \log k} \limsup_{n\to\infty} \mathrm{I}(\pi^{(\ell)}_{S,n};\pi^{(\ell)}_{T,n})$$

respectively.

Definition. The *lower* and *upper block mutual information rates between* $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ are

$$\mathrm{I}(S;T) = \lim_{\ell\to\infty} \mathrm{I}_\ell(S;T)$$

and

$$\hat{\mathrm{I}}(S;T) = \lim_{\ell\to\infty} \hat{\mathrm{I}}_\ell(S;T),$$

respectively.

We now present the main theorem of this section, which states that the lower and upper block mutual information rates coincide with the lower and upper finite-state mutual dimensions, respectively.

Theorem 11. *For all* $S, T \in \Sigma^\infty$*,*

$$mdim_{FS}(S:T) = \mathrm{I}(S;T)$$

and

$$Mdim_{FS}(S:T) = \hat{\mathrm{I}}(S;T).$$

The final result of this section regarding the properties of block-mutual information rates between sequences follows directly from Theorem 8, Theorem 9, Corollary 10, and Theorem 11.

Theorem 12 (Properties of Block Mutual Information Rates)**.** *For all* $S, T \in \Sigma^\infty$*,*

1) $\mathrm{H}(S)+\mathrm{H}(T)-\hat{\mathrm{H}}(S,T)\leq\mathrm{I}(S;T)\leq\hat{\mathrm{H}}(S)+\hat{\mathrm{H}}(T)-\hat{\mathrm{H}}(S,T)$.
2) $\mathrm{H}(S)+\mathrm{H}(T)-\mathrm{H}(S,T)\leq\hat{\mathrm{I}}(S;T)\leq\hat{\mathrm{H}}(S)+\hat{\mathrm{H}}(T)-\mathrm{H}(S,T)$.
3) $\mathrm{I}(S;T)\leq\min\{\mathrm{H}(S),\mathrm{H}(T)\}, \hat{\mathrm{I}}(S;T)\leq\min\{\hat{\mathrm{H}}(S),\hat{\mathrm{H}}(T)\}$.
4) $0 \leq \mathrm{I}(S;T) \leq \hat{\mathrm{I}}(S;T) \leq 1$.
5) $\mathrm{I}(S;S) = \mathrm{H}(S)$, $\hat{\mathrm{I}}(S;S) = \hat{\mathrm{H}}(S)$.
6) $\mathrm{I}(S;T) = \mathrm{I}(T;S)$, $\hat{\mathrm{I}}(S;T) = \hat{\mathrm{I}}(T;S)$.

V. Finite-State Mutual Dimension and Independence

In this section we explore some of the relationships between finite-state mutual dimension and normal sequences. More specifically, we provide necessary and sufficient conditions for when two normal sequences achieve finite-state mutual dimension zero.

Becher, Carton, and Heiber provided a notion of *finite-state independence* using the *conditional compression ratio* of a sequence *given* another sequence. Specifically, they define two sequences $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ to be *finite-state independent* if the conditional compression ratio $\rho(S\,|\,T)$ of S given T is equal to the compression ratio $\rho(S)$ of S, the conditional compression ratio $\rho(T\,|\,S)$ of T given S is equal to the compression ratio $\rho(T)$ of T, and both $\rho(S)$ and $\rho(T)$ are greater than zero. In their investigation they showed that, for any two normal sequences $R_1 \in \Sigma^\infty$ and $R_2 \in \Sigma^\infty$, if R_1 and R_2 are finite-state independent, then (R_1, R_2) is normal. However, they also showed that the converse does not hold, i.e., there are two normal sequences R_1 and R_2 such that (R_1, R_2) is normal and not finite-state independent [3]. Alvarez, Becher, and Carton also proved several characterizations of finite-state independence using various kinds of Büchi automata [1].

We now proceed to discuss the concept of normality and its relationship to finite-state dimension. A probability measure α on Σ is *positive* if, for every $a \in \Sigma$, $\alpha(a) > 0$. A *probability measure* on Σ^∞ is a function $\nu : \Sigma^* \to [0,1]$ with the following two properties.

1) $\nu(\lambda) = 1$.
2) For all $w \in \Sigma^*, \nu(w) = \sum_{a\in\Sigma} \nu(wa)$.

Intuitively, $\nu(w)$ is the probability that $w \sqsubseteq S$ (w is a prefix of S) when $S \in \Sigma^\infty$ is "chosen according to" the probability measure ν. Every probability measure α on Σ induces the probability measure α on Σ^∞ defined by

$$\alpha(w) = \prod_{i=1}^{|w|} \alpha(w[i]),$$

for all $w \in \Sigma^*$.

Definition. Let α be a probability measure on Σ, $S \in \Sigma^\infty$, and $\ell \in \mathbb{Z}^+$.

1) S is *α-ℓ-normal* if, for all $x \in \Sigma^\ell$,
$$\lim_{n\to\infty} \pi_{S,n}(x) = \alpha(x).$$
2) S is *α-normal* if S is α-ℓ-normal for all $\ell \in \mathbb{Z}^+$.
3) S is *normal* if S is μ-normal, where μ is the uniform probability measure on Σ.

Lutz proved the following theorem about α-normal sequences [16].

Theorem 13 ([16])**.** *If α is a probability measure on Σ, then, for every α-normal sequence $R \in \Sigma^\infty$,*

$$dim_{FS}(R) = Dim_{FS}(R) = \frac{\mathrm{H}(\alpha)}{\log k}.$$

Our first theorem of this section is a "mutual" version of Theorem 13.

Theorem 14. *If α is a probability measure on $\Sigma \times \Sigma$, then, for every α-normal sequence $(R_1, R_2) \in (\Sigma\times\Sigma)^\infty$,*

$$mdim_{FS}(R_1:R_2) = Mdim_{FS}(R_1:R_2) = \frac{\mathrm{I}(\alpha_1;\alpha_2)}{\log k}.$$

Schnorr and Stimm proved a characterization of normal sequences in terms of finite-state gamblers [21]. Later, Dai, Lathrop, Lutz, and Mayordomo showed that any normal sequence achieves finite-state dimension one [7], while Bourke, Hitchcock, and Vinodchandran showed that any sequence that achieves finite-state dimension one is normal [4]. Therefore, a sequence is normal if and only if $dim_{FS}(S) = 1$.

The main theorem of this section provides a similar characterization for pairs of normal sequences that achieve finite-state mutual dimension zero. Note that, in this theorem, the *product probability measure* $(\alpha_1 \times \alpha_2)$ on $\Sigma \times \Sigma$ is defined by

$$(\alpha_1 \times \alpha_2)(a, b) = \alpha_1(a)\alpha_2(b),$$

for all $a, b \in \Sigma$.

Theorem 15. *Let α_1 and α_2 be positive probability measures on Σ. If $R_1 \in \Sigma^\infty$ is α_1-normal and $R_2 \in \Sigma^\infty$ is α_2-normal, then (R_1, R_2) is $(\alpha_1 \times \alpha_2)$-normal if and only if $Mdim_{FS}(R_1 : R_2) = 0$.*

When $\alpha_1 = \alpha_2 = \mu$ in the above theorem, we obtain the following corollary.

Corollary 16. *For all normal sequences $R_1, R_2 \in \Sigma^\infty$, (R_1, R_2) is normal if and only if $Mdim_{FS}(R_1 : R_2) = 0$.*

Thus finite-state mutual dimension provides a mechanism in which to reason about the degree to which two sequences are independent of one another at the finite-state level.

References

[1] Nicolás Alvarez, Verónica Becher, and Oliver Carton. Finite-state independence and normal sequences. *Journal of Computer and System Sciences*, 103:1–17, 2019.

[2] Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, and Elvira Mayordomo. Effective strong dimension in algorithmic information and computational complexity. *SIAM Journal of Computing*, 37(3):671–705, 2007.

[3] Verónica Becher, Oliver Carton, and Pablo Ariel Heiber. Finite-state independence. *Theory of Computing Systems*, 62:1555–1572, 2018.

[4] C. Bourke, J.M. Hitchcock, and N.V. Vinodchandran. Entropy rates and finite-state dimension. *Theoretical Computer Science*, 3:392–406, 2005.

[5] Adam Case and Jack H. Lutz. Mutual dimension. *ACM Transactions on Computation Theory*, 7:1–26, 2015.

[6] Adam Case and Jack H. Lutz. Mutual dimension and random sequences. *Theoretical Computer Science*, 731:68–97, 2018.

[7] Jack J. Dai, James I. Lathrop, Jack H. Lutz, and Elvira Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310:1–33, 2004.

[8] Randall Dougherty, Jack H. Lutz, Daniel R. Mauldin, and Jason Teutsch. Translating the Cantor set by a random real. *Transactions of the American Mathematical Society*, 366(6):3027–3041, 2014.

[9] Émile Borel. Sur les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909.

[10] Xiaoyang Gu, Jack H. Lutz, Elvira Mayordomo, and Philippe Moser. Dimension spectra of random subfractals of self-similar fractals. *Annals of Pure and Applied Logic*, 165:1707–1726, 2014.

[11] John M. Hitchcock. Correspondence principles for effective dimensions. *Theory of Computing Systems*, 38(5):559–571, 2005.

[12] Xiang Huang, Jack H. Lutz, Elvira Mayordomo, and Donald M. Stull. Asymptotic divergences and strong dichotomy. *IEEE Transactions on Information Theory*, 67(10):6296–6305, 2021.

[13] Alexander Kozachinskiy and Alexander Shen. Two Characterizations of Finite-State Dimension. In *Fundamentals of Computation Theory. FCT 2019. Lecture Notes in Computer Science*, volume 11651. Springer, 2019.

[14] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York, fourth edition, 2008.

[15] Jack H. Lutz. The dimension of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.

[16] Jack H. Lutz. A divergence formula for randomness and dimension. *Theoretical Computer Science*, 412(1-2):166–177, January 2011.

[17] Jack H. Lutz and Neil Lutz. Algorithmic information, plane Kakeya sets, and conditional dimension. *ACM Transactions on Computation Theory*, 10:1–22, 2018.

[18] Jack H. Lutz, Neil Lutz, and Elvira Mayordomo. Extending the reach of the point-to-set principle. In *The Proceedings of the 39th International Symposium on Theoretical Aspects of Computer Science*, pages 48:1–48:14, Marseille, France, 2022. Schloss Dagstuhl.

[19] Jack H. Lutz and Elvira Mayordomo. Dimensions of points in self-similar fractals. *SIAM Journal on Computing*, 38(3):1080–1112, 2008.

[20] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Information Processing Letters*, 84(1):1–3, 2002.

[21] C.P. Schnorr and H. Stimm. Endliche automaten und zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

[22] Dafna Sheinwald. On the Ziv-Lempel Proof and Related Topics. *Proceedings of the IEEE*, 82(6):866–871, 1994.

[23] Jacob Ziv and Abraham Lempel. Compression of Individual Sequences via Variable-Rate Coding. *IEEE Transactions on Information Theory*, 24(5):530–536, 1978.