Finite-State Mutual Dimension*

Adam Case† Jack H. Lutz†‡
Drake University Iowa State University

Roughly speaking, a *finite-state compressor* (FSC) can be thought of as a *finite-state automata* that does not have any accepting states and is equipped with an output function. An FSC C processes an input string $u \in \Sigma^*$ by reading u from left to right, transitioning to a new state and outputting a string after each symbol of u is read. The *output* of C on an input $u \in \Sigma^*$, denoted by $C(u)$, is the concatenation of all the output strings produced while reading the input string. An *information-lossless finite-state compressor* (ILFSC) C is a FSC whose input string $u \in \Sigma^*$ may be recovered when given knowledge of the output string $C(u)$ and the last state that C entered upon reading u.

The *k-state compression ratio of* $u \in \Sigma^*$ is

$$\rho_k(u) = \min\left\{ \frac{|C(u)|}{|u| \log |\Sigma|} \,\middle|\, C \text{ is an ILFSC on } \Sigma \text{ that has } k \text{ states} \right\}.$$

The *lower* and *upper finite-state dimensions* of a sequence $S \in \Sigma^\infty$ are

$$dim_{FS}(S) = \lim_{k\to\infty} \liminf_{n\to\infty} \rho_k(S \upharpoonright n) \quad \text{and} \quad Dim_{FS}(S) = \lim_{k\to\infty} \limsup_{n\to\infty} \rho_k(S \upharpoonright n),$$

respectively, and are also known as the *lower* and *upper finite-state compression ratios* of a sequence S [5, 4, 1]. Intuitively, these quantities represent the lower and upper *densities* of finite-state information contained within a sequence.

In this talk, we introduce "mutual" versions of finite-state dimension. More specifically, we define the *lower* and *upper finite-state mutual dimensions* between $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ by

$$mdim_{FS}(S : T) = \lim_{r\to\infty} \lim_{t\to\infty} \liminf_{n\to\infty} \left[\rho_t(S \upharpoonright n) + \rho_t(T \upharpoonright n) - \rho_r(S \upharpoonright n, T \upharpoonright n)\right]$$

and

$$Mdim_{FS}(S : T) = \lim_{r\to\infty} \lim_{t\to\infty} \limsup_{n\to\infty} \left[\rho_t(S \upharpoonright n) + \rho_t(T \upharpoonright n) - \rho_r(S \upharpoonright n, T \upharpoonright n)\right],$$

*A preprint of the article that this abstract is based on may be found on arXiv:2109.14574 [3].

†Part of this research was carried out while the authors participated in the program "Equidistribution: Arithmetic, Computational and Probabilistic Aspects" at the National University of Singapore Institute for Mathematical Sciences in 2019.

‡This research was supported in part by National Science Foundation grants 1545028 and 1900716.

respectively, and they can be thought of as the lower and upper *densities* of finite-state information *shared* by the sequences S and T. In our first result, we show that these definitions have all of the properties expected of a measure of mutual information.

We relate finite-state mutual dimension to entropy rates of sequences [5]. For any $\ell \in \mathbb{Z}^+$, $x \in \Sigma^\ell$, and sequence $S \in \Sigma^\infty$, the *frequency* of x in the first $n \cdot \ell$ symbols of S is denoted by $\pi^{(\ell)}_{S,n}(x)$, and we can demonstrate that $\pi^{(\ell)}_{S,n}$ is a discrete probability measure on the set of strings Σ^ℓ of length ℓ. We define the *lower* and *upper block mutual information rates* between $S \in \Sigma^\infty$ and $T \in \Sigma^\infty$ by

$$\mathrm{I}(S;T) = \lim_{\ell\to\infty} \frac{1}{\ell \log|\Sigma|} \liminf_{n\to\infty} \mathrm{I}(\pi^{(\ell)}_{S,n};\pi^{(\ell)}_{T,n}) \quad \text{and} \quad \hat{\mathrm{I}}(S;T) = \lim_{\ell\to\infty} \frac{1}{\ell \log|\Sigma|} \limsup_{n\to\infty} \mathrm{I}(\pi^{(\ell)}_{S,n};\pi^{(\ell)}_{T,n}),$$

respectively, where $\mathrm{I}(\pi^{(\ell)}_{S,n};\pi^{(\ell)}_{T,n})$ is the *Shannon mutual information* between the discrete probability measures $\pi^{(\ell)}_{S,n}$ and $\pi^{(\ell)}_{T,n}$. In our next result, we show that finite-state mutual dimension may be characterized in terms of block mutual information rates. That is, we show that, for all $S, T \in \Sigma^\infty$, $mdim_{FS}(S:T) = \mathrm{I}(S;T)$ and $Mdim_{FS}(S:T) = \hat{\mathrm{I}}(S;T)$.

There have been several interesting interactions between the study of finite-state dimension and the concept of *normality*. We say that a sequence $R \in \Sigma^\infty$ is *normal* if every string of the same length occurs with the same limiting frequency within R, and it has been shown that a sequence $R \in \Sigma^\infty$ is normal if and only if $dim_{FS}(R) = 1$ [2]. In our final result, we provide a similar characterization for pairs of normal sequences that achieve finite-state mutual dimension zero. More specifically, we show that if $R_1 \in \Sigma^\infty$ and $R_2 \in \Sigma^\infty$ are normal sequences, then $(R_1, R_2) \in (\Sigma \times \Sigma)^\infty$ is normal if and only if $Mdim_{FS}(R_1 : R_2) = 0$, where (R_1, R_2) is the sequence obtained by pairing each symbol from R_1 with the symbol from R_2 at the same index.

References

[1] 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.

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

[3] Adam Case and Jack H. Lutz. Finite-state mutual dimension. arXiv:2109.14574, 2021.

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

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