Finite-State Dimension and Real Arithmetic

David Doty [*], Jack H. Lutz [†], and Satyadev Nandakumar [‡]

Department of Computer Science, Iowa State University, Ames, IA 50011, USA.
{ddoty,lutz,satyadev} *at* cs *dot* iastate *dot* edu

Abstract. We use entropy rates and Schur concavity to prove that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α, the base-k expansions of $\alpha, q + \alpha$, and $q\alpha$ all have the same finite-state dimension and the same finite-state strong dimension. This extends, and gives a new proof of, Wall's 1949 theorem stating that the sum or product of a nonzero rational number and a Borel normal number is always Borel normal.

1 Introduction

The finite-state dimension of a sequence S over a finite alphabet Σ is an asymptotic measure of the density of information in S as perceived by finite-state automata. This quantity, denoted $\dim_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical Hausdorff dimension [15, 12] introduced by Dai, Lathrop, Lutz, and Mayordomo [9]. A dual quantity, the finite-state *strong* dimension of S, denoted $\mathrm{Dim}_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical packing dimension [30, 29, 12] introduced by Athreya, Hitchcock, Lutz, and Mayordomo [2]. (Explicit definitions of $\dim_{\mathrm{FS}}(S)$ and $\mathrm{Dim}_{\mathrm{FS}}(S)$ appear in section 2.) In fact both $\dim_{\mathrm{FS}}(S)$ and $\mathrm{Dim}_{\mathrm{FS}}(S)$ are asymptotic measures of the density of finite-state information in S, with $0 \leq \dim_{\mathrm{FS}}(S) \leq \mathrm{Dim}_{\mathrm{FS}}(S) \leq 1$ holding in general. The identity $\dim_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(S)$ holds when S is sufficiently "regular", but, for *any* two real numbers $0 \leq \alpha \leq \beta \leq 1$, there exists a sequence S with $\dim_{\mathrm{FS}}(S) = \alpha$ and $\mathrm{Dim}_{\mathrm{FS}}(S) = \beta$ [13].

Although finite-state dimension and finite-state strong dimension were originally defined in terms of finite-state gamblers [9, 2] (following the gambling approach used in the first effectivizations of classical fractal dimension [21, 22]), they have also been shown to admit equivalent definitions in terms of information-lossless finite-state compressors [9, 2], finite-state predictors in the log-loss model [16, 2], and block-entropy rates [6]. In each case, the definitions of $\dim_{\mathrm{FS}}(S)$ and $\mathrm{Dim}_{\mathrm{FS}}(S)$ are exactly dual, differing only that a limit inferior appears in one definition where a limit superior appears in the other. These

two finite-state dimensions are thus, like their counterparts in fractal geometry, robust quantities and not artifacts of a particular definition.

The sequences S satisfying $\dim_{\mathrm{FS}}(S) = 1$ are precisely the *(Borel) normal* sequences, i.e., those sequences in which each nonempty string $w \in \Sigma^*$ appears with limiting frequency $|\Sigma|^{-|w|}$. (This fact was implicit in the work of Schnorr and Stimm [27] and pointed out explicitly in [6].) The normal sequences, introduced by Borel in 1909 [4], were extensively investigated in the twentieth century [25, 19, 32, 10, 14]. Intuitively, the normal sequences are those sequences that are random relative to finite-state automata. This statement may seem objectionable when one first learns that the Champernowne sequence

$$01000110110000010100111100\ldots,$$

obtained by concatenating all binary strings in standard order, is normal [8], but it should be noted that a finite-state automaton scanning this sequence will spend nearly all its time in the middle of long strings that are random in the (stronger) sense of Kolmogorov complexity [20] and, having only finite memory, will have no way of "knowing" where such strings begin or end. This perspective is especially appropriate when modeling situations in which a data stream is truly massive relative to the computational resources of the entity processing it.

An informative line of research on normal sequences concerns operations that preserve normality. For example, in his 1949 Ph.D. thesis under D.H. Lehmer, Wall [31] proved that every subsequence that is selected from a normal sequence by taking all symbols at positions occurring in a given arithmetical progression is itself normal. Agafonov [1] extended this by showing that every subsequence of a normal sequence that is selected using a regular language is itself normal; Kamae [17] and Kamae and Weiss [18] proved related results; and Merkle and Reimann [24] proved that a subsequence selected from a normal sequence using a context-free language need not be normal (in fact, can be constant, even if selected by a one-counter language). For another example, again in his thesis, Wall [31] (see also [19, 5]) proved that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α that is normal base k (i.e., has a base-k expansion that is a normal sequence), the sum $q + \alpha$ and the product $q\alpha$ are also normal base k. (It should be noted that a real number α may be normal in one base but not in another [7, 26].)

This paper initiates the study of operations that preserve finite-state dimension and finite-state strong dimension. This study is related to, but distinct from, the study of operations that preserve normality. It is clear that every operation that preserves finite-state dimension must also preserve normality, but the converse does not hold. For example, a subsequence selected from a sequence according an arithmetical progression need not have the same finite-state dimension as the original sequence. This is because a sequence with finite-state dimension less than 1 may have its information content distributed heterogeneously. Specifically, given a normal sequence S over the alphabet $\{0, 1\}$, define a sequence T whose n^{th} bit is the $\frac{n}{2}^{\mathrm{th}}$ bit of S if n is even and 0 otherwise. Then the sequence S and the constant sequence 0^{∞} are both selected from T

according to arithmetic progressions, but it is easy to verify that $\dim_{\mathrm{FS}}(T) = \mathrm{Dim}_{\mathrm{FS}}(T) = \frac{1}{2}, \dim_{\mathrm{FS}}(0^{\infty}) = \mathrm{Dim}_{\mathrm{FS}}(0^{\infty}) = 0$, and $\dim_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(S) = 1$. Hence, Wall's first above-mentioned theorem does not extend to the preservation of finite-state dimension. Of course, this holds *a fortiori* for the stronger results by Agafonov, Kamae, and Weiss.

Our main theorem states that Wall's second above-mentioned theorem, unlike the first one, does extend to the preservation of finite-state dimension. That is, we prove that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α, the base-k expansions of $\alpha, q + \alpha$, and $q\alpha$ all have the same finite-state dimension and the same finite-state strong dimension.

The proof of our main theorem does not, and probably cannot, resemble Wall's uniform distribution argument. Instead we use Bourke, Hitchcock, and Vinodchandran's block-entropy rate characterizations of $\dim_{\mathrm{FS}}$ and $\mathrm{Dim}_{\mathrm{FS}}$ [6], coupled with the Schur concavity of the entropy function [28, 23, 3], to prove that finite-state dimension and finite-state strong dimension are contractive functions with respect to a certain "logarithmic block dispersion" pseudometric that we define on the set of all infinite k-ary sequences. (A function is contractive if the distance between its values at sequences S and T is no more than the pseudodistance between S and T.) This gives a general method for bounding the difference between the finite-state dimensions, and the finite-state strong dimensions, of two sequences. We then use this method to prove our main theorem. In particular, this gives a new proof of Wall's theorem on the sums and products of rational numbers with normal numbers.

In summary, our main result is a fundamental theorem on finite-state dimension that is a quantitative extension of a classical theorem on normal numbers but requires a different, more powerful proof technique than the classical theorem.

2   Preliminaries

Throughout this paper, $\Sigma = \{0, 1, \ldots, k - 1\}$, where $k \geq 2$ is an integer. All *strings* are elements of Σ^*, and all *sequences* are elements of Σ^{∞}. If x is a string or sequence and i, j are integers, $x[i \mathinner{.\,.} j]$ denotes the string consisting of the ith through j^{th} symbols in x, provided that these symbols exist. We write $x[i] = x[i \mathinner{.\,.} i]$ for the ith symbol in x, noting that $x[0]$ is the leftmost symbol in x. If w is a string and x is a string or sequence, we write $w \sqsubseteq x$ to indicate that $w = x[0 \mathinner{.\,.} n - 1]$ for some nonnegative integer n.

A *base-k expansion* of a real number $\alpha \in [0, 1]$ is a sequence $S \in \Sigma^{\infty}$ such that

$$\alpha = \sum_{n=0}^{\infty} S[n] k^{-(n+1)}.$$

A sequence $S \in \Sigma^{\infty}$ is *(Borel) normal* if, for every nonempty string $w \in \Sigma^+$

$$\lim_{n \to \infty} \frac{1}{n} \left| \left\{ u \in \Sigma^{<n} \,\middle|\, uw \sqsubseteq S \right\} \right| = |\Sigma|^{-|w|},$$

i.e., if each string w appears with asymptotic frequency $k^{-|w|}$ in S.

If Ω is a nonempty finite set, we write $\Delta(\Omega)$ for the set of all (discrete) probability measures on Ω, i.e., all functions $\pi : \Omega \to [0, 1]$ satisfying $\sum\limits_{w \in \Omega} \pi(w)$ $= 1$. We write $\Delta_n = \Delta(\{1, \ldots, n\})$.

All logarithms in this paper are base 2. The *Shannon entropy* of a probability measure $\pi \in \Delta(\Omega)$ is

$$H(\pi) = \sum_{w \in \Omega} \pi(w) \log \frac{1}{\pi(w)},$$

where $0 \log \frac{1}{0} = 0$.

We briefly define finite-state dimension and finite-state strong dimension. As noted in the introduction, several equivalent definitions of these dimensions are now known. In this paper, it is most convenient to use the definitions in terms of block-entropy rates, keeping in mind that Bourke, Hitchcock, and Vinodchandran [6] proved that these definitions are equivalent to earlier ones.

For nonempty strings $w, x \in \Sigma^+$, we write

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

for the number of *block occurrences* of w in x. Note that $0 \leq \#_\square(w, x) \leq \frac{|x|}{|w|}$.

For each sequence $S \in \Sigma^\infty$, positive integer n, and string $w \in \Sigma^{<n}$, the nth *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\square(w, S[0 \ldots n|w| - 1])}{n}.$$

Note that, for all $S \in \Sigma^\infty$ and $0 < l < n$,

$$\sum_{w \in \Sigma^l} \pi_{S,n}(w) = 1,$$

i.e., $\pi_{S,n}^{(l)} \in \Delta(\Sigma^l)$, where we write $\pi_{S,n}^{(l)}$ for the restriction of $\pi_{S,n}$ to Σ^l.

For each sequence $S \in \Sigma^\infty$ and positive integer l, the lth *normalized lower and upper block entropy rates* of S are

$$H_l^-(S) = \frac{1}{l \log k} \liminf_{n \to \infty} H\left(\pi_{S,n}^{(l)}\right)$$

and

$$H_l^+(S) = \frac{1}{l \log k} \limsup_{n \to \infty} H\left(\pi_{S,n}^{(l)}\right),$$

respectively.

Definition 1. *Let* $S \in \Sigma^\infty$.

1. *The* finite-state dimension *of S is*

$$\dim_{\text{FS}}(S) = \inf_{l \in \mathbb{Z}^+} H_l^-(S).$$

2. The finite-state strong dimension *of S is*

$$\mathrm{Dim}_{\mathrm{FS}}(S) = \inf_{l \in \mathbb{Z}^+} H_l^+(S).$$

More discussion and properties of these dimensions appear in the references cited in the introduction, but this material is not needed to follow the technical arguments in the present paper.

3 Logarithmic Dispersion and Finite-State Dimension

In this section we prove a general theorem stating that the difference between two sequences' finite-state dimensions (or finite-state strong dimensions) is bounded by a certain "pseudodistance" between the sequences. Recall that $\Delta_n = \Delta(\{1, \ldots, n\})$ is the set of all probability measures on $\{1, \ldots, n\}$.

Definition 2. *Let n be a positive integer. The* logarithmic dispersion *(briefly, the* log-dispersion*) between two probability measures* $\pi, \mu \in \Delta_n$ *is*

$$\delta(\pi, \mu) = \log m,$$

where m is the least positive integer for which there is an $n \times n$ nonnegative real matrix $A = (a_{ij})$ with the following three properties.

    (i) A is stochastic*: each column of A sums to 1, i.e., $\sum_{i=1}^n a_{ij} = 1$ holds for all $1 \leq j \leq n$.*
    (ii) $A\pi = \mu$, i.e., $\sum_{j=1}^n a_{ij}\pi(j) = \mu(i)$ holds for all $1 \leq i \leq n$.
    (iii) No row or column of A contains more than m nonzero entries.

It is clear that $\delta : \Delta_n \times \Delta_n \to [0, \log n]$. We now extend δ to a normalized function $\delta^+ : \Sigma^\infty \times \Sigma^\infty \to [0, 1]$. Recall the block-frequency functions $\pi_{S,n}^{(l)}$ defined in section 2.

Definition 3. *The* normalized upper logarithmic block dispersion *between two sequences $S, T \in \Sigma^\infty$ is*

$$\delta^+(S, T) = \limsup_{l \to \infty} \frac{1}{l \log k} \limsup_{n \to \infty} \delta\left(\pi_{S,n}^{(l)}, \pi_{T,n}^{(l)}\right).$$

Recall that a *pseudometric* on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying the following three conditions for all $x, y, z \in X$.

    (i) $d(x, y) \geq 0$, with equality if $x = y$.   (nonnegativity)
    (ii) $d(x, y) = d(y, x)$.   (symmetry)
    (iii) $d(x, z) \leq d(x, y) + d(y, z)$.   (triangle inequality)

(A pseudometric is a *metric*, or *distance function*, on X if it satisfies (i) with "if" replaced by "if and only if".) The following fact must be known, but we do not know a reference at the time of this writing.

Lemma 3.1. *For each positive integer n, the log-dispersion function δ is a pseudometric on Δ_n.*

It is easy to see that S is not a metric on Δ_n for any $n \geq 2$. For example, if π is any nonuniform probability measure on $\{1, \ldots, n\}$ and μ obtained from π by permuting the values of π nontrivially, then $\pi \neq \mu$ but $\delta(\pi, \mu) = 0$.

Lemma 3.1 has the following immediate consequence.

Corollary 3.2. *The normalized upper log-block dispersion function δ^+ is a pseudometric on Σ^∞.*

If d is a pseudometric on a set X, then a function $f : X \to \mathbb{R}$ is *d-contractive* if, for all $x, y \in X$,

$$|f(x) - f(y)| \leq d(x, y),$$

i.e., the distance between $f(x)$ and $f(y)$ does not exceed the pseudodistance between x and y. We use Schur concavity to prove the following lemma in the Technical Appendix.

Lemma 3.3. *For each positive integer n, the Shannon entropy function $H : \Delta_n \to [0, \log n]$ is δ-contractive.*

The following useful fact follows easily from Lemma 3.3.

Theorem 3.4. *Finite-state dimension and finite-state strong dimension are δ^+-contractive. That is, for all $S, T \in \Sigma^\infty$,*

$$|\dim_{\mathrm{FS}}(S) - \dim_{\mathrm{FS}}(T)| \leq \delta^+(S, T)$$

and

$$|\mathrm{Dim}_{\mathrm{FS}}(S) - \mathrm{Dim}_{\mathrm{FS}}(T)| \leq \delta^+(S, T).$$

In this paper, we only use the following special case of Theorem 3.4.

Corollary 3.5. *Let $S, T \in \Sigma^\infty$. If*

$$\limsup_{n \to \infty} \delta\left(\pi_{S,n}^{(l)}, \pi_{T,n}^{(l)}\right) = o(l)$$

as $l \to \infty$, then

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

and

$$\mathrm{Dim}_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(T).$$

4 Finite-State Dimension and Real Arithmetic

Our main theorem concerns real numbers rather than sequences, so the following notation is convenient. For each real number α and each integer $k \geq 2$, write

$$\dim_{\mathrm{FS}}^{(k)}(\alpha) = \dim_{\mathrm{FS}}(S)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha) = \mathrm{Dim}_{\mathrm{FS}}(S),$$

where S is a base-k expansion of $\alpha - \lfloor \alpha \rfloor$. Note that this notation is well-defined, because a real number α has two base-k expansions if and only if it is a k-adic rational, in which case both expansions are eventually periodic and hence have finite-state strong dimension 0. It is routine to verify the following.

Observation 4.1. *For every integer $k \geq 2$, every positive integer m, and every real number α,*

$$\dim_{\mathrm{FS}}^{(k)}(m + \alpha) = \dim_{\mathrm{FS}}^{(k)}(-\alpha) = \dim_{\mathrm{FS}}^{(k)}(\alpha)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(m + \alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(-\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha).$$

The following lemma contains most of the technical content of our main theorem.

Lemma 4.2 (main lemma). *For every integer $k \geq 2$, every positive integer m, and every real number $\alpha \geq 0$,*

$$\dim_{\mathrm{FS}}^{(k)}(m\alpha) = \dim_{\mathrm{FS}}^{(k)}(\alpha)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(m\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha).$$

Proof. Let k, m, and α be as given, let $S, T \in \Sigma^\infty$ be the base-k expansions of $\alpha - \lfloor \alpha \rfloor$, $m\alpha - \lfloor m\alpha \rfloor$, respectively, and write

$$\pi_{\alpha,n}^{(l)} = \pi_{S,n}^{(l)} \quad , \quad \pi_{m\alpha,n}^{(l)} = \pi_{T,n}^{(l)}$$

for each $l, n \in \mathbb{Z}^+$. By Corollary 3.5, it suffices to show that

$$\limsup_{n \to \infty} \delta\left(\pi_{\alpha,n}^{(l)}, \pi_{m\alpha,n}^{(l)}\right) = o(l) \tag{4.1}$$

as $l \to \infty$.

Let $r = \lfloor \log_k m \rfloor$, let

$$m = \sum_{i=0}^{r} m_i k^i$$

be the base-k expansion of m, and let

$$s = \sum_{i=0}^{r} m_i.$$

The first thing to note is that, in base k, $m\alpha - \lfloor m\alpha \rfloor$ is the sum, modulo 1, of s copies of $\alpha - \lfloor \alpha \rfloor$, with m_i of these copies shifted i symbols to the left, for each $0 \le i \le r$.

For each $l \in \mathbb{Z}^+$ and $j \in \mathbb{N}$, let

$$u_j^{(l)} = S[jl \mathinner{.\,.} (j+1)l - 1],$$
$$v_j^{(l)} = T[jl \mathinner{.\,.} (j+1)l - 1]$$

be the jth l-symbol blocks of $\alpha - \lfloor \alpha \rfloor$, $m\alpha - \lfloor m\alpha \rfloor$, respectively. If we let

$$\tau_j^{(l)} = \sum_{i=0}^{r} m_i \sum_{t=(j+1)l}^{\infty} S[t+i]k^{-(t+1)}$$

be the sum of the tails of the above-mentioned s copies of $\alpha - \lfloor \alpha \rfloor$ lying to the right of the jth l-symbol block, then the block $v_j^{(l)}$ of $m\alpha - \lfloor m\alpha \rfloor$ is completely determined by $u_j^{(l)}$, the "carry"

$$c_j^{(l)} = \left\lfloor k^{(j+1)l} \tau_j^{(l)} \right\rfloor,$$

and the longest string of symbols shifted from the right, which is the string $u_{j+1}^{(l)}[0 \mathinner{.\,.} r-1]$. To be more explicit, note that

$$0 \le c_j^{(l)} \le k^{(j+1)l} \tau_j^{(l)} \le k^{(j+1)l} \sum_{i=0}^{r} m_i \sum_{t=(j+1)l}^{\infty} (k-1)k^{-(t+1)} = s;$$

define the "advice"

$$h_j^{(l)} = \left(c_j^{(l)}, u_j^{(l)}[0 \mathinner{.\,.} r-1] \right) \in \{0, \dots, s\} \times \Sigma^r;$$

and define the function

$$f^{(l)} : \Sigma^l \times \{0, \dots, s\} \times \Sigma^r \to \Sigma^l$$

by letting $f^{(l)}(x, c, z)$ be the base-k expansion of the integer

$$mn_x^{(k)} + c + \sum_{i=0}^{r} m_i \sum_{t=0}^{i-1} z[t]k^t \ \textbf{mod} \ k^l,$$

where $n_x^{(k)}$ is the nonnegative integer of which x is a base-k expansion, possibly with leading 0's. (Intuitively, the three terms here are the "block product", the "carry", and the "shift", respectively.) Then, for all integers $l > 0$ and $j \ge 0$,

$$v_j^{(l)} = f^{(l)}(u_j^{(l)}, h_j^{(l)}).$$

For positive integers l and n, define the $k^l \times k^l$ matrix $A^{(l,n)} = \left(a_{y,x}^{(l,n)} \right)$ by

$$a_{y,x}^{(l,n)} = \begin{cases} \dfrac{\left| \left\{ j < n \ \middle| \ u_j^{(l)} = x \text{ and } f^{(l)}\left(x, h_j^{(l)}(j)\right) = y \right\} \right|}{n\pi_{\alpha,n}^{(l)}(x)} & \text{if } \pi_{\alpha,n}^{(l)}(x) > 0 \\ 1 & \text{if } \pi_{\alpha,n}^{(l)}(x) = 0 \text{ and } x = y \\ 0 & \text{otherwise} \end{cases}$$

for all $x, y \in \Sigma^l$. It is routine to verify that

$$\sum_{y \in \Sigma^l} a_{y,x}^{(l,n)} = 1$$

for all $x \in \Sigma^l$, i.e., $A^{(l,n)}$ is stochastic, and that

$$\sum_{x \in \Sigma^l} a_{y,x}^{(l,n)} \pi_{\alpha,n}^{(l)}(x) = \pi_{m\alpha,n}^{(l)}(y)$$

for all $y \in \Sigma^l$, i.e., $A^{(l,n)} \pi_{\alpha,n}^{(l)} = \pi_{m\alpha,n}^{(l)}$. We complete the proof by bounding the number of nonzero entries in each row and column of $A^{(l,n)}$.

Fix a column x of $A^{(l,n)}$. If $\pi_{\alpha,n}^{(l)}(x) = 0$, then there is exactly one nonzero entry in column x of $A^{(l,n)}$. If $\pi_{\alpha,n}^{(l)}(x) > 0$, then the number of nonzero entries in column x is bounded by

$$|\{0, \ldots, s\} \times \Sigma^r| = (s+1)k^r \le (s+1)m.$$

Hence there are at most $(s+1)m$ nonzero entries in column x of $A^{(l,n)}$.

Fix a row y of $A^{(l,n)}$. Let g be the greatest common divisor of m and k^l. Note that, for all $n_1, n_2 \in \mathbb{Z}^+$,

$$mn_1 \equiv mn_2 \mod k^l \implies k^l \mid m(n_2 - n_1)$$
$$\implies \frac{k^l}{g} \ \middle|\ \frac{m}{g}(n_2 - n_1)$$
$$\implies \frac{k^l}{g} \ \middle|\ n_2 - n_1$$
$$\implies n_1 \equiv n_2 \mod \frac{k^l}{g}.$$

This implies that each string $y \in \Sigma^l$ has at most g preimages x under the mapping that takes x to the base-k expansion of $mn_x^{(l)} \mod k^l$. This, in turn, implies that there are at most $g|\{0, \ldots, s\} \times \Sigma^r| \le g(s+1)m$ nonzero entries in row y of $A^{(l,n)}$.

We have shown that, for each $l, n \in \mathbb{Z}^+$, the matrix $A^{(l,n)}$ testifies that

$$\delta\left(\pi_{\alpha,n}^{(l)}, \pi_{m\alpha,n}^{(l)}\right) \le \log(g(s+1)m) \le \log(m^2(s+1)).$$

Since this bound does not depend on l or n, this proves (4.1). □

We now prove that addition and multiplication by nonzero rationals preserve finite-state dimension and finite-state strong dimension.

Theorem 4.3 (main theorem). *For every integer $k \geq 2$, every nonzero rational number q, and every real number α,*

$$\dim_{\text{FS}}^{(k)}(q + \alpha) = \dim_{\text{FS}}^{(k)}(q\alpha) = \dim_{\text{FS}}^{(k)}(\alpha)$$

and

$$\text{Dim}_{\text{FS}}^{(k)}(q + \alpha) = \text{Dim}_{\text{FS}}^{(k)}(q\alpha) = \text{Dim}_{\text{FS}}^{(k)}(\alpha).$$

Proof. Let k, q, and α be as given, and write $q = \frac{a}{b}$, where a and b are integers with $a \neq 0$ and $b > 0$. By Observation 4.1 and Lemma 4.2,

$$\begin{aligned}
\dim_{\text{FS}}^{(k)}(q\alpha) &= \dim_{\text{FS}}^{(k)}\left(\frac{|a|}{b}\alpha\right) &= \dim_{\text{FS}}^{(k)}\left(b\frac{|a|}{b}\alpha\right) \\
&= \dim_{\text{FS}}^{(k)}(|a|\alpha) &= \dim_{\text{FS}}^{(k)}(\alpha),
\end{aligned}$$

and

$$\begin{aligned}
\dim_{\text{FS}}^{(k)}(q + \alpha) &= \dim_{\text{FS}}^{(k)}\left(\frac{a}{b} + \alpha\right) &= \dim_{\text{FS}}^{(k)}\left(\frac{a + b\alpha}{b}\right) \\
&= \dim_{\text{FS}}^{(k)}\left(b\frac{a + b\alpha}{b}\right) &= \dim_{\text{FS}}^{(k)}(a + b\alpha) \\
&= \dim_{\text{FS}}^{(k)}(b\alpha) &= \dim_{\text{FS}}^{(k)}(\alpha).
\end{aligned}$$

Similarly, $\text{Dim}_{\text{FS}}^{(k)}(q\alpha) = \text{Dim}_{\text{FS}}^{(k)}(\alpha)$, and $\text{Dim}_{\text{FS}}^{(k)}(q + \alpha) = \text{Dim}_{\text{FS}}^{(k)}(\alpha)$. □

Finally, we note that Theorem 4.3 gives a new proof of the following classical theorem.

Corollary 4.4. *(Wall [31]) Let $k \geq 2$. For every nonzero rational number q and every real number α that is normal base k, the sum $q + \alpha$ and the product $q\alpha$ are also normal base k.*

References

1. V. N. Agafonov. Normal sequences and finite automata. *Soviet Mathematics Doklady*, 9:324–325, 1968.
2. K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information, and computational complexity. *SIAM Journal on Computing*. To appear.
3. R. Bhatia. *Matrix Analysis*. Springer, 1997.
4. E. Borel. Sur les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909.

5. J. Borwein and D. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century.* A. K. Peters, Ltd., Natick, MA, 2004.

6. C. Bourke, J. M. Hitchcock, and N. V. Vinodchandran. Entropy rates and finite-state dimension. *Theoretical Computer Science*, 349:392–406, 2005.

7. J. W. S. Cassels. On a problem of Steinhaus about normal numbers. *Colloquium Mathematicum*, 7:95–101, 1959.

8. D. G. Champernowne. Construction of decimals normal in the scale of ten. *J. London Math. Soc.*, 2(8):254–260, 1933.

9. J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310:1–33, 2004.

10. K. Dajani and C. Kraaikamp. *Ergodic Theory of Numbers.* The Mathematical Association of America, 2002.

11. G. A. Edgar. *Classics on Fractals.* Westview Press, Oxford, U.K., 2004.

12. K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley & Sons, 1990.

13. X. Gu, J. H. Lutz, and P. Moser. Dimensions of Copeland-Erdös sequences. *Information and Computation*, 2005.

14. G. Harman. One hundred years of normal numbers. In M. A. Bennett, B. C. Berndt, N. Boston, H. G. Diamond, A. J. Hildebrand, and W. Philip (eds.)*, Surveys in Number Theory: Papers from the Millennial Conference on Number Theory*, pages 57–74, 2003.

15. F. Hausdorff. Dimension und äusseres Mass. *Mathematische Annalen*, 79:157–179, 1919. English version appears in [11], pp. 75-99.

16. J. M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1–3):431–441, 2003.

17. T. Kamae. Subsequences of normal sequences. *Israel Journal of Mathematics*, 16:121–149, 1973.

18. T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*, 21:101–110, 1975.

19. L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences.* Wiley-Interscience, 1974.

20. M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, Berlin, 1997. Second Edition.

21. J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32:1236–1259, 2003.

22. J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003.

23. A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York, 1979.

24. W. Merkle and J. Reimann. On selection functions that do not preserve normality. In B. Rovan and P. Vojtás, editors, *MFCS*, volume 2747 of *Lecture Notes in Computer Science*, pages 602–611, Bratislava, Slovakia, 2003. Springer.

25. I. Niven. *Irrational Numbers.* Wiley, 1956.

26. W. Schmidt. On normal numbers. *Pacific Journal of Mathematics*, 10:661–672, 1960.

27. C. P. Schnorr and H. Stimm. Endliche Automaten und Zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

28. I. Schur. Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

29. D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Mathematica*, 153:259–277, 1984.

30. C. Tricot. Two definitions of fractional dimension. *Mathematical Proceedings of the Cambridge Philosophical Society*, 91:57–74, 1982.

31. D. D. Wall. *Normal Numbers*. PhD thesis, University of California, Berkeley, California, USA, 1949.

32. B. Weiss. *Single Orbit Dynamics*. American Mathematical Society, Providence, RI, 2000.

5 Technical Appendix

This appendix contains proofs of Lemmas 3.1 and 3.3.

Proof (of Lemma 3.1). Let n be a positive integer and let $\pi, \mu, \nu \in \Delta_n$. Since $\delta(\pi, \mu) = \log m$, where m is a positive integer, $\delta(\pi, \mu) \geq 0$. Thus δ is nonnegative. If $\pi = \mu$, then it is easy to verify that the $n \times n$ identity matrix I_n testifies that $\delta(\pi, \mu) = 0$.

   To show that δ is symmetric, it suffices to prove that $\delta(\pi, \mu) \leq \delta(\mu, \pi)$. Let $m = 2^{\delta(\mu, \pi)}$. Then there exists an $n \times n$ nonnegative stochastic matrix $A = (a_{ij})$ such that $\pi = A\mu$ and A has at most m nonzero entries in each row and column. Define the $n \times n$ matrix $A' = (a'_{ij})$ for all $1 \leq i, j \leq n$ by

$$
a'_{ij} = \begin{cases} a_{ji}\dfrac{\mu(i)}{\pi(j)}, & \text{if } \pi(j) > 0 \\ a_{ji}\dfrac{1}{\sum_{k=1}^{n} a_{jk}}, & \text{if } \pi(j) = 0 \end{cases}
$$

For all $1 \leq j \leq n$ such that $\pi(j) = 0$,

$$
\sum_{i=1}^{n} a'_{ij} = \sum_{i=1}^{n} a_{ji}\frac{1}{\sum_{k=1}^{n} a_{jk}} = 1.
$$

Since $A\mu = \pi$, for all $1 \leq j \leq n$ such that $\pi(j) > 0$,

$$
\sum_{i=1}^{n} a'_{ij} = \sum_{i=1}^{n} a_{ji}\frac{\mu(i)}{\pi(j)} = \frac{1}{\pi(j)} \sum_{i=1}^{n} a_{ji}\mu(i) = \frac{1}{\pi(j)}\pi(j) = 1,
$$

so A' is stochastic. Since A is stochastic, for all $1 \leq i \leq n$,

$$
\sum_{j=1}^{n} a'_{ij}\pi(j) = \sum_{j=1}^{n} \left(a_{ji}\frac{\mu(i)}{\pi(j)} \right) \pi(j) = \sum_{j=1}^{n} a_{ji}\mu(i) = \mu(i),
$$

so $A'\pi = \mu$. Since $a_{ji} = 0 \implies a'_{ij} = 0$, and A has at most m nonzero entries in each row and column, A' has at most m nonzero entries in each row and column as well. Thus $\delta(\pi, \mu) \leq \log m = \delta(\mu, \pi)$, so δ is symmetric.

   To see that δ satisfies the triangle inequality, let $m_1 = 2^{\delta(\pi, \mu)}$ and $m_2 = 2^{\delta(\mu, \nu)}$. It suffices to show that $\delta(\pi, \nu) \leq \log m_1 + \log m_2 = \log m_1 m_2$. There exist $n \times n$ nonnegative stochastic matrices A_1 and A_2 having no more than m_1 and m_2 nonzero entries in each row and column, respectively, satisfying $A_1\pi = \mu$ and $A_2\mu = \nu$. Let $A = A_2 A_1$. Since the product of two stochastic matrices is stochastic, A is stochastic. Also, $A\pi = A_2(A_1\pi) = A_2\mu = \nu$. Finally, since no row or column of A_1 (resp. A_2) contains more than m_1 (resp. m_2) nonzero entries, no row or column of A contains more than $m_1 m_2$ nonzero entries. Thus $\delta(\pi, \nu) \leq \log m_1 m_2$, so δ satisfies the triangle inequality. $\square$

The proof of Lemma 3.3 uses Schur concavity [28, 23, 3], which we now review. We say that a vector $\overrightarrow{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is *nonincreasing* if $x_1 \geq \ldots \geq x_n$. If $\overrightarrow{x}, \overrightarrow{y} \in \mathbb{R}^n$ are nonincreasing, then we say that $\overrightarrow{x}$ *majorizes* $\overrightarrow{y}$, and we write $\overrightarrow{x} \succcurlyeq \overrightarrow{y}$, if the following two conditions hold.

(i) $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.
(ii) For all $1 \leq t \leq n$, $\sum_{i=1}^{t} x_i \geq \sum_{i=1}^{t} y_i$.

Given a vector $\overrightarrow{x} \in \mathbb{R}^n$ and a permutation π of $\{1, \ldots, n\}$, write $\pi(\overrightarrow{x}) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Call a set $D \subseteq \mathbb{R}^n$ *symmetric* if $\pi(\overrightarrow{x}) \in D$ holds for every $\overrightarrow{x} \in D$ and every permutation π of $\{1, \ldots, n\}$. For $D \subseteq \mathbb{R}^n$, a function $f : D \to \mathbb{R}$ is then *symmetric* if D is symmetric and $f(\overrightarrow{x}) = f(\pi(\overrightarrow{x}))$ holds for every $\overrightarrow{x} \in D$ and every permutation π of $\{1, \ldots, n\}$.

Definition 4. *Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$ be symmetric. Then f is* Schur-concave *if, for all $\overrightarrow{x}, \overrightarrow{y} \in \mathbb{R}^n$,*

$$\overrightarrow{x} \succcurlyeq \overrightarrow{y} \implies f(\overrightarrow{x}) \leq f(\overrightarrow{y}).$$

The set Δ_n of all probability measures on $\{1, \ldots, n\}$ can be regarded as the $(n-1)$-dimensional simplex

$$\Delta_n = \left\{ \overrightarrow{p} \in [0,1]^n \ \middle| \ \sum_{i=1}^{n} p_i = 1 \right\} \subseteq \mathbb{R}^n.$$

This set Δ_n is symmetric, as is the Shannon entropy function $H : \Delta_n \to [0, \log n]$. In fact, the following fundamental property of Shannon entropy is well known [3].

Lemma 5.1. *The Shannon entropy function $H : \Delta_n \to [0, \log n]$ is Schur-concave.*

We now use Lemma 5.1 to prove Lemma 3.3.

Proof (of Lemma 3.3). Fix a positive integer n, and let $\overrightarrow{p}, \overrightarrow{q} \in \Delta_n$. By the symmetry of δ (established in Lemma 3.1), it suffices to prove that

$$H(\overrightarrow{p}) \leq H(\overrightarrow{q}) + \delta(\overrightarrow{p}, \overrightarrow{q}). \tag{5.1}$$

Without loss of generality, assume that $\overrightarrow{p}$ and $\overrightarrow{q}$ are nonincreasing. Let m be the positive integer such that $\delta(\overrightarrow{p}, \overrightarrow{q}) = \log m$, and let $A = (a_{ij})$ be an $n \times n$ matrix testifying to the value of $\delta(\overrightarrow{p}, \overrightarrow{q})$. Define an $n \times n$ matrix $B = (b_{ij})$ by

$$b_{ij} = \begin{cases} 1 & \text{if } (i-1)m < j \leq \min\{im, n\}; \\ 0 & \text{otherwise.} \end{cases}$$

That is, the first block of m entries in the first row of B are 1's, the second block of m entries in the second row of B are 1's, and so on, until the last $n - m\left(\left\lceil \frac{n}{m} \right\rceil - 1\right)$ entries in the $\left\lceil \frac{n}{m} \right\rceil^{\text{th}}$ row of B are 1's.

Let $\overrightarrow{r} = B\overrightarrow{p}$. Intuitively, B represents the "worst-case" matrix with no more than m nonzero entries in each row and column, in the sense that it produces the vector with the lowest entropy. More formally, we show that $\overrightarrow{r}$ majorizes the vector $\overrightarrow{q}$, and thus $\overrightarrow{r}$ has entropy at most that of $\overrightarrow{q}$. However, since B is limited to m nonzero entries in each row and column, it cannot redistribute the values in $\overrightarrow{p}$ by too much, so the entropy of $\overrightarrow{r}$ will be close to that of $\overrightarrow{p}$.

Since B is stochastic (because each column contains exactly one 1) and $\overrightarrow{p} \in \Delta_n$, we have $\overrightarrow{r} \in \Delta_n$. Clearly, $\overrightarrow{r}$ is nonincreasing. For each $1 \leq j \leq n$, let $C_i = \{j \mid a_{ij} > 0\}$, noting that $|C_i| \leq m$. Then, for all $1 \leq t \leq n$,

$$
\begin{aligned}
\sum_{i=1}^{t} r_i &= \sum_{i=1}^{t}\sum_{j=1}^{n} b_{ij}p_j = \sum_{i=1}^{t}\sum_{j=(i-1)m+1}^{\min\{im,n\}} p_j = \sum_{i=1}^{\min\{tm,n\}} p_i \\
&\geq \sum_{j \in C_1 \cup \ldots \cup C_t} p_j \geq \sum_{i=1}^{t}\sum_{j \in C_i} a_{ij}p_j = \sum_{i=1}^{t}\sum_{j=1}^{n} a_{ij}p_j = \sum_{i=1}^{t} q_i.
\end{aligned}
$$

(The first inequality holds because $\overrightarrow{p}$ is nonincreasing and each $|C_i| \leq m$. The second inequality holds because $\sum_{i=1}^{n} a_{ij} = 1$ holds for each $1 \leq j \leq n$, whence a single p_j's appearances in various C_i's collectively contribute at most p_j to the sum on the right.) This shows that $\overrightarrow{r} \succcurlyeq \overrightarrow{q}$, whence Lemma 5.1 tells us that $H(\overrightarrow{r}) \leq H(\overrightarrow{q})$. It follows by Jensen's inequality and the (ordinary) concavity of the logarithm that

$$
\begin{aligned}
H(\overrightarrow{p}) &\leq H(\overrightarrow{p}) + \log \sum_{i=1}^{n} p_i \frac{1}{p_i} r_{\lceil \frac{i}{m} \rceil} - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m} \rceil} \right) \\
&= H(\overrightarrow{p}) + \log \sum_{i=1}^{n} r_{\lceil \frac{i}{m} \rceil} - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m} \rceil} \right) \\
&\leq H(\overrightarrow{p}) + \log m - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m} \rceil} \right) \\
&= \sum_{i=1}^{n} p_i \log \frac{1}{r_{\lceil \frac{i}{m} \rceil}} + \log m \\
&= \sum_{i=1}^{n} r_i \log \frac{1}{r_i} + \log m \\
&= H(\overrightarrow{r}) + \delta(\overrightarrow{p}, \overrightarrow{q}) \\
&\leq H(\overrightarrow{q}) + \delta(\overrightarrow{p}, \overrightarrow{q}),
\end{aligned}
$$

i.e., (5.1) holds. $\qquad\square$