Normal Numbers and Finite Automata

Verónica Becher
vbecher@dc.uba.ar

Pablo Ariel Heiber
pheiber@dc.uba.ar

Departamento de Computación, Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires & CONICET, Argentina

August 17, 2012

Abstract

We give an elementary and direct proof of the following theorem: A real number is normal to a given integer base if, and only if, its expansion in that base is incompressible by lossless finite-state compressors (these are finite automata augmented with an output transition function such that the automata input-output behaviour is injective; they are also known as injective finite-state transducers). As a corollary we obtain V.N. Agafonov's theorem on the preservation of normality on subsequences selected by finite automata, generalized to arbitrary alphabets.

1 Statement and Discussion of Results

In this note we give an elementary and direct proof of the following:

Characterization Theorem. *A real number is normal to a given integer base if, and only if, its expansion expressed in that base is incompressible by lossless finite-state compressors.*

Normality, defined by Émile Borel in 1909 [2], requires that the infinite expansion of a real number be evenly balanced: a real number is normal to a given integer base if every block of digits of the same length occurs with the same limit frequency in the expansion of the number expressed in that base. For example, if a number is normal to base two, each of the digits '0' and '1' occur, in the limit, half of the times; each of the blocks '00', '01', '10' and '11' occur one fourth of the times, and so on. *Lossless finite-state compressors*, introduced by David Huffman in 1959 [8], are ordinary finite automata augmented with an output transition function such that the automata input-output behaviour is injective. They are also called injective finite-state transducers.

Although the Characterization Theorem has not hitherto appeared explicitly in print, it was known to the experts in the field as a consequence of these two results:

(a) Schnorr and Stimm in 1971 [12] considered *martingales* constructed from finite automata increased with stationary transition probabilities and used them to predict the symbols in a sequence. They proved that normal sequences are exactly those at which no such martingale succeeds in making unbounded profit. The proof relies on the theory of Markov chains.*

(b) Dai, Lathrop, Lutz and Mayordomo in 2004 [6] defined *finite-state dimension* as a measure of how much success is achievable by the martingales considered by Schnorr and Stimm. Bounded success corresponds to finite-state dimension one. Their theorem establishes that the finite-state dimension of a sequence is the infimum of all compression ratios achievable on the sequence by lossless finite-state compressors. Therefore, finite-state dimension one is equivalent to incompressibility by lossless finite-state compressors.

In [6] the authors also showed that every sequence normal to base two has finite-state dimension one (the result generalizes to any other base). Bourke, Hitchcock and Vinodchandran [3]

*For instance, Lemma 2 in [12] uses that in a Markov chain with discrete time, the sequence of positive probabilities in recurrent states has a Cesàro limit.

established the converse using the notion of *entropy rate* for blocks of symbols. These results together amount to yet another proof of Schnorr and Stimm's theorem: normality coincides with finite-state dimension one.

Our proof of the Characterization Theorem shows the incompressibility of normal numbers bypassing the intermediate property of finite-state dimension one. The proof is done directly in terms of finite automata, with elementary counting arguments and basic concepts in the theory of prefix codes.

As a corollary we obtain a a theorem due to V.N. Agafonov in 1968 on the preservation of normality on subsequences selected by finite automata [1].

Agafonov's Theorem. *Let $\mathcal{A}$ be the binary alphabet. An infinite sequence is normal to the alphabet $\mathcal{A}$ if, and only if, every infinite subsequence selected by a finite automaton is, again, normal to alphabet $\mathcal{A}$.*

Agafonov's publication [1] does not include the complete proof (it depends on previous work only available in the Russian literature). M.O'Connor [11] gave it using predictors defined from finite automata, and Broglio and Liardet [4] generalized it to arbitrary alphabets. We also obtain Agafonov's theorem for arbitrary alphabets.

It is known that for some slightly more powerful automata Agafonov's theorem fails: Merkle and Reimann [10] showed that normality is not preserved in subsequences selected by deterministic one-counter automata (pushdown automata with a unary stack alphabet) nor by linear languages (languajes recognized by one-turn pushdown automata, namely, the automata with limited operations on one stack: once they start popping, they must stop pushing).

Whether finite automata is the largest class that yields normality-preserving selectors is yet to be determined. Similarly, the largest class of machines that can not compress normal numbers remains to be known.

2 Basic Definitions

Hereafter $\mathcal{A}$ and $\mathcal{B}$ are alphabets (finite sets of at least two symbols), $\mathcal{A}^n$ is the set of strings of n symbols from $\mathcal{A}$, $\mathcal{A}^{<n} = \bigcup_{i=0}^{n-1} \mathcal{A}^i$ is the set of strings of length strictly less than n, $\mathcal{A}^*$ is the set of finite strings of any length and $\mathcal{A}^\omega$ is the set of infinite sequences of symbols from $\mathcal{A}$. $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$ and observe that $|\mathcal{A}^n| = |\mathcal{A}|^n$. λ is the empty string, $|s|$ is the length of string s, $s[i]$ is the symbol at position i of s, for $1 \leq i \leq |s|$, and $s[i..i+k-1]$ is the string of k consecutive symbols of s starting at position i, for $1 \leq i \leq |s| - k + 1$. We use a similar notation for the infinite sequences in $\mathcal{A}^\omega$.

2.1 Normal Numbers

There are several equivalent definitions of normality[†]; we give here the one that is most convenient to prove the Characterization Theorem. For notational purposes we present it directly on infinite sequences of symbols from an alphabet $\mathcal{A}$.

Definition. A sequence $\alpha \in \mathcal{A}^\omega$ is *simply normal* to alphabet $\mathcal{A}$ if each individual symbol in $\mathcal{A}$ has the same asymptotic frequency in α,

$$\forall x \in \mathcal{A}, \quad \lim_{k\to\infty} \frac{\mathrm{occ}(x, \alpha[1..k])}{k} = \frac{1}{|\mathcal{A}|},$$

[†]Borel's original definition, given in [2], says: A real number r is *simply normal* to a given integer base b if each digit in $\{0, 1, ..., b-1\}$ has the same asymptotic frequency $1/b$ in the expansion of r expressed in base b. A real number r is normal to base b if each of the numbers r, br, b^2r.. are simply normal to the bases b^n, for every $n \geq 1$. Although it seems more demanding, this last condition is equivalent to require that just r be simply normal to the bases b^n, for every $n \geq 1$. Another equivalent definition is in terms of equifrequency of *blocks of digits*, for every block size. An alternative characterization proves that a real number x is normal to a base b if and only if, the sequence $(xb^n)_{n\geq 1}$ is uniformly distributed modulo one. For a proof of these equivalences see, for instance, [5].

where $\mathrm{occ}(x, s) = |\{i : x = s[i]\}|$ is the number of occurrences of the symbol x in string s. A sequence $\alpha \in \mathcal{A}^\omega$ is *normal* to alphabet $\mathcal{A}$ if it is simply normal to alphabet $\mathcal{A}^n$, for every $n \geq 1$.

Thus, normality to a given alphabet implies normality to any power of that alphabet. And if a sequence is normal to a given base, then so is each of its final segments.

2.2 Finite-State Compressors

A finite-state compressor is a finite automaton with two tapes, an input tape and an output tape.

Definition. A *finite-state compressor* is a 6-tuple $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o \rangle$ where $\mathcal{A}$ is the input alphabet, $\mathcal{B}$ is the output alphabet, $\mathcal{Q}$ is a finite set of states, $q_0 \in \mathcal{Q}$ is the initial state, $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is the transition function and $o : \mathcal{Q} \times \mathcal{A} \to \mathcal{B}^*$ is the output function. The automaton processes the input symbols according to the current state q. When a symbol $x \in \mathcal{A}$ is read, the automaton moves to state $\delta(q, x)$ and outputs $o(q, x)$. We extend δ and o to process strings:

$$\begin{array}{lcllcl} \delta^*(q, \lambda) & = & q, & o^*(q, \lambda) & = & \lambda, \\ \delta^*(q, xs) & = & \delta^*(\delta(q, x), s). & o^*(q, xs) & = & o(q, x)o^*(\delta(q, x), s). \end{array}$$

We write $C(s)$ for $o^*(q_0, s)$ and $|C(s)|$ for its length.

Definition. A finite-state compressor is *lossless* if, from a given output and finishing state, there is at most one input that produces it from the initial state. This is equivalent to requiring that the function $f(s) = \langle o^*(q_0, s), \delta^*(q_0, s) \rangle$ be injective.

Definition. The *compression ratio* for a finite-state compressor C of a string $s \in \mathcal{A}^*$ is the output length divided by the length of a standard optimal coding of s in symbols of alphabet $\mathcal{B}$:

$$\rho_C(s) = \frac{|C(s)|}{|s| \ \log_{|\mathcal{B}|} |\mathcal{A}|}.$$

The compression ratio for a finite-state compressor C of an infinite sequence $\alpha \in \mathcal{A}^\omega$ is

$$\rho_C(\alpha) = \liminf_{n \to \infty} \rho_C(\alpha[1..n]).$$

The finite-state compression ratio of a given sequence α is the infimum of the compression ratios achievable by all finite-state compressors, namely,

$$\rho(\alpha) = \inf\{\rho_C(\alpha) : C \text{ is a lossless finite-state compressor}\}.$$

If C is a lossless finite-state compressor, then any string s can be coded by a sequence of length $|C(s)| + k_C$, where k_C depends on C itself and its finishing state, but not on s. Since the constant does not significantly affect the compression ratio for sufficiently long strings, it makes sense to refer to $|C(s)|$ as the length of the compression of s.

Definition. A finite-state compressor C *compresses* a sequence $\alpha \in \mathcal{A}^\omega$ if the compression ratio for C of α, $\rho_C(\alpha)$, is strictly less than one. A sequence α is *compressible* by lossless finite-state compressors if there is a lossless finite-state compressor C that compresses α, or equivalently, if $\rho(\alpha) < 1$.

The next lemma proves that any given sequence in alphabet $\mathcal{A}$ and the same sequence seen in the alphabet power $\mathcal{A}^n$, for any given natural number n, have the same finite-state compression ratio.

Lemma 1. *Fix $n \in \mathbb{N}$. Let α be a sequence in $\mathcal{A}^\omega$ and let α^n be the sequence in $(\mathcal{A}^n)^\omega$ such that $\alpha^n[i] = \alpha[n(i-1)+1...ni]$. Then, α and α^n have the same finite-state compression ratio.*

Proof. Given a lossless finite-state compressor C with input alphabet $\mathcal{A}$ we can construct another one with input alphabet $\mathcal{A}^n$ that gives the same output. And, conversely, given a compressor with input alphabet $\mathcal{A}^n$ we can construct another that reads symbols from $\mathcal{A}$.

To go from $\mathcal{A}$ to $\mathcal{A}^n$, we combine the processing of n consecutive symbols into one transition. Let $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o\rangle$ be a finite-state compressor. For a given n, we define $C^n = \langle \mathcal{A}^n, \mathcal{B}, \mathcal{Q}, q_0, \delta^*, o^*\rangle$ by restricting the domain of δ^* and o^* to $\mathcal{A}^n$. Notice that $(\mathcal{A}^n)^* \subseteq \mathcal{A}^*$, $(\delta^*)^* = \delta^*$ and $(o^*)^* = o^*$, so $|C^n(s)| = |C(s)|$ for strings s in $(\mathcal{A}^n)^*$.

To go from $\mathcal{A}^n$ to $\mathcal{A}$ we reverse the previous conversion: a single transition that processes a symbol in $\mathcal{A}^n$ is split into n individual transitions, so this gives rise to new intermediate states. In the new automaton we need a different set of states, that we define by considering the prefix tree of strings in $\mathcal{A}$ of length less than n. Given a finite-state compressor $C^n = \langle \mathcal{A}^n, \mathcal{B}, \mathcal{Q}, q_0, \delta, o\rangle$, we define $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q} \times \mathcal{A}^{<n}, \langle q_0, \lambda\rangle, \delta', o'\rangle$, where

$$\delta'(\langle q, s\rangle, x) = \begin{cases} \langle q, sx\rangle & ,\text{if } |s| < n-1, \\ \langle \delta(q, sx), \lambda\rangle & ,\text{if } |s| = n-1. \end{cases} \qquad o'(\langle q, s\rangle, x) = \begin{cases} \lambda & ,\text{if } |s| < n-1, \\ o(q, sx) & ,\text{if } |s| = n-1. \end{cases}$$

The fact that $(\delta')^*(\langle q, \lambda\rangle, s) = \langle \delta^*(q, s), \lambda\rangle$ and $(o')^*(\langle q, \lambda\rangle, s) = o^*(q, s)$ follows directly by applying n times the definitions above. □

Observe that in the proof above, if one starts from a given automaton and applies the two transformations successively, one does not recover the same automaton (because one of the transformations changes the set of states and the other does not). Although the two automata are different, their outputs coincide. It is possible, with some extra work, to perform both transformations and recover the original automaton: If we start from an automaton with input alphabet $\mathcal{A}^n$, apply the transformation to alphabet $\mathcal{A}$, remove the unreachable states, apply the other transformation and then rename the remaining states. If we start with an automaton with input alphabet $\mathcal{A}$, after the two transformations use a minimization algorithm to unify equivalent states.

3 Proof of the Characterization Theorem

3.1 Normal Implies Incompressible

Assume $\alpha \in \mathcal{A}^\omega$ is normal to alphabet $\mathcal{A}$. Let $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o\rangle$ be an arbitrary lossless finite-state compressor such that all its states are reachable and let $\varepsilon > 0$ be an arbitrarily small real. Using elementary counting arguments, we show that the compression ratio for C of α is strictly larger than $(1-\varepsilon)^3$.

First, for each string $s \in \mathcal{A}^*$, let a_s be the minimum addition to the output length that could result from processing s:

$$a_s = \min\{|o^*(q, s)| : q \in \mathcal{Q}\}.$$

For each length n, consider the set of strings in $\mathcal{A}^n$ whose processing yields a large contribution to the output length:

$$\mathcal{S}_n = \{s \in \mathcal{A}^n : a_s > (1-\varepsilon) n \log_{|\mathcal{B}|} |\mathcal{A}|\}.$$

Given a string s in the complement set $\mathcal{A}^n \setminus \mathcal{S}_n$ there is at least one processing of s that produces an output of length at most $(1-\varepsilon) n \log_{|\mathcal{B}|} |\mathcal{A}|$. However, C is lossless and each state is reachable, so the string s can be associated in a unique way to a starting state q_s, an ending state $\delta^*(q_s, s)$, and an output string $o^*(q_s, s)$ of length at most $(1-\varepsilon) n \log_{|\mathcal{B}|} |\mathcal{A}|$. Consequently, the assignment $f(s) = \langle q_s, \delta^*(q_s, s), o^*(q_s, s)\rangle$, for an appropriate q_s defines an injective function

$$f : \mathcal{A}^n \setminus \mathcal{S}_n \to \mathcal{Q} \times \mathcal{Q} \times \mathcal{B}^{<(1-\varepsilon) n \log_{|\mathcal{B}|} |\mathcal{A}|+1}.$$

Thus, we can bound the cardinals of $\mathcal{A}^n \setminus \mathcal{S}_n$ and $\mathcal{S}_n$ as follows:

$$\begin{array}{lcll} |\mathcal{A}^n \setminus \mathcal{S}_n| & < & |\mathcal{Q}|^2 |\mathcal{B}|^{(1-\varepsilon) n \log_{|\mathcal{B}|} |\mathcal{A}|+1} & = & |\mathcal{B}|\ |\mathcal{Q}|^2\ |\mathcal{A}|^{(1-\varepsilon) n}. \\ |\mathcal{S}_n| & > & |\mathcal{A}^n| - |\mathcal{B}||\mathcal{Q}|^2 |\mathcal{A}|^{(1-\varepsilon) n} & = & |\mathcal{A}|^n (1 - |\mathcal{B}||\mathcal{Q}|^2 |\mathcal{A}|^{-\varepsilon n}). \end{array}$$

Since $|\mathcal{B}||\mathcal{Q}|^2|\mathcal{A}|^{-\varepsilon n}$ is arbitrarily close to 0 for a sufficiently large n, let n be large enough such that $|\mathcal{S}_n| > |\mathcal{A}|^n(1-\varepsilon)$.

Let C^n be the transformation of compressor C by changing its input alphabet from $\mathcal{A}$ to $\mathcal{A}^n$, as in the proof of Lemma 1. Let $\alpha^n \in (\mathcal{A}^n)^\omega$ be the sequence α seen in alphabet $\mathcal{A}^n$, that is, $\alpha^n[i] = \alpha[(i-1)n+1..in]$. By the definition of normality, since α is normal to alphabet $\mathcal{A}$, α^n is simply normal to alphabet $\mathcal{A}^n$. Then, let k_0 be such that

$$\forall k > k_0 \;\; \forall x \in \mathcal{A}^n \qquad \frac{\mathrm{occ}(x, \alpha^n[1..k])}{k} \;>\; |\mathcal{A}^n|^{-1}(1-\varepsilon) \;=\; |\mathcal{A}|^{-n}(1-\varepsilon).$$

To give a lower bound for the compression length of C^n on $\alpha^n[1..k]$, for $k > k_0$, we consider only the strings $s \in \mathcal{S}_n$ yielding a large contribution to the output length. For each such s, we sum up the length of the output produced by each occurrence of s in $\alpha^n[1..k]$:

$$\begin{aligned}
|C^n(\alpha^n[1..k])| &\geq \sum_{x\in\mathcal{A}^n} \mathrm{occ}(x, \alpha^n[1..k])\; a_x \\
&> \sum_{x\in\mathcal{S}_n} k\;|\mathcal{A}|^{-n}(1-\varepsilon)\; a_x \\
&> \sum_{x\in\mathcal{S}_n} k\;|\mathcal{A}|^{-n}(1-\varepsilon)(1-\varepsilon)(n\log_{|\mathcal{B}|}|\mathcal{A}|) = |\mathcal{S}_n|k|\mathcal{A}|^{-n}(1-\varepsilon)^2\log_{|\mathcal{B}|}|\mathcal{A}^n| \\
&> (1-\varepsilon)^3 k\;\log_{|\mathcal{B}|}|\mathcal{A}^n|.
\end{aligned}$$

Therefore, the compression ratio for C^n of α^n,

$$\rho_{C^n}(\alpha^n) = \liminf_{k\to\infty} \frac{|C^n(\alpha^n[1..k])|}{k\;\log_{|\mathcal{B}|}|\mathcal{A}^n|}$$

is at least $(1-\varepsilon)^3$. By the invariance of the compression ratio under transformations to powers of the alphabet, proved in Lemma 1, $\rho_C(\alpha)$ is also at least $(1-\varepsilon)^3$.

3.2 Not Normal Implies Compressible

Assume $\alpha \in \mathcal{A}^\omega$ is not normal to alphabet $\mathcal{A}$. We will show that α is compressible (regardless of the choice of an output alphabet $\mathcal{B}$). Since α is not normal to alphabet $\mathcal{A}$, there is some n such that α is not simply normal to alphabet $\mathcal{A}^n$. Fix such a block length n. As before, let α^n be the sequence with symbols in $\mathcal{A}^n$ such that $\alpha^n[i] = \alpha[(i-1)n+1..in]$. Then there is some $x \in \mathcal{A}^n$ such that

$$\lim_{k\to\infty} \frac{\mathrm{occ}(x, \alpha^n[1..k])}{k} \neq \frac{1}{|\mathcal{A}^n|}.$$

Either this limit does not exist, or it is different from $|\mathcal{A}|^{-n}$. Thus, it is impossible that both,

$$\liminf_{k\to\infty} \frac{\mathrm{occ}(x, \alpha^n[1..k])}{k} \quad \text{and} \quad \limsup_{k\to\infty} \frac{\mathrm{occ}(x, \alpha^n[1..k])}{k},$$

be equal to $1/|\mathcal{A}^n|$. We will define an increasing sequence of positions $(i_k)_{k\in\mathbb{N}}$ relative to this block length n such that for each $y \in A^n$, the limiting frequency of y at positions $(i_k)_{k\in\mathbb{N}}$,

$$f_y = \lim_{k\to\infty} \frac{\mathrm{occ}(x, \alpha^n[1..i_k])}{i_k}$$

is defined and $f_x \neq |A^n|^{-1}$. Let $y_1 = x$ and for $j = 2, .., |\mathcal{A}^n|$, let y_j be the j-th element of $\mathcal{A}^n \setminus \{x\}$ in the lexicographic order. We define $(i_k)_{k\in\mathbb{N}}$ by taking subsequences. Let $\left(i_k^{(1)}\right)_{k\in\mathbb{N}}$ be an increasing sequence of positions such that f_{y_1} is defined and different from $|A^n|^{-1}$. This exists because $y_1 = x$ and we already argued that the limit for x is not $|A^n|^{-1}$ over all subsequences.

And for each j, $2 \leq j \leq |\mathcal{A}|^n$, let $\left(i_k^{(j)}\right)_{k\in\mathbb{N}}$ be a subsequence of $\left(i_k^{(j-1)}\right)_{k\in\mathbb{N}}$ such that the limit f_{y_j} is defined when considered over positions $\left(i_k^{(j)}\right)$. Since frequencies are bounded between 0 and 1, such a subsequence necessarily exists. Observe that the sequence $\left(i_k^{(|\mathcal{A}^n|)}\right)_{k\in\mathbb{N}}$ verifies that for all $y \in \mathcal{A}^n$, f_y is defined. By letting $i_k = i_k^{(|\mathcal{A}^n|)}$, for each $k \in \mathbb{N}$, we obtain the desired sequence.

We now prove that α^n is compressible. We shall bound the compression ratio of α^n at the sequence of positions $(i_k)_{k\in\mathbb{N}}$. We follow an idea known from information theory as in the Noiseless-Coding Theorem [13]. We encode the blocks via a block-to-variable-length encoding, with m-length blocks of symbols from A^n, such that the average codeword-length is less than $m \times n$. For each integer m, we define a compressor C_m that codes the input by groups of m symbols. Let $C_m = \langle \mathcal{A}^n, \mathcal{B}, (\mathcal{A}^n)^{<m}, \lambda, \delta, o\rangle$ be such that for each $q \in (\mathcal{A}^n)^{<m}$ and $z \in \mathcal{A}^n$,

$$\delta(q,z) = \begin{cases} qz & ,\text{if } |q| < m-1, \\ \lambda & ,\text{if } |q| = m-1. \end{cases} \qquad o(q,z) = \begin{cases} \lambda & ,\text{if } |q| < m-1, \\ \bar{o}(qz) & ,\text{if } |q| = m-1, \end{cases}$$

where $\bar{o} : (\mathcal{A}^n)^m \to \mathcal{B}^*$ is an injective map into a prefix-free subset of $\mathcal{B}^*$ such that

$$|\bar{o}(z_1 z_2 ... z_m)| = \left\lceil \sum_{i=1}^{m} -\log_{|\mathcal{B}|} f_{z_i} \right\rceil.$$

Since

$$\sum_{s\in(\mathcal{A}^n)^m} |\mathcal{B}|^{-|\bar{o}(s)|} = \sum_{s\in(\mathcal{A}^n)^m} \prod_{j=1}^{m} f_{s[j]} \leq 1,$$

such a prefix-free set exists (for example, it can be defined by Huffman's coding [7]). This ensures that C_m is lossless. We now give an upper bound for the length of the output of C_m on an arbitrary string $s \in (\mathcal{A}^n)^*$. Fix p to be the largest integer such that $pm \leq |s|$. By definition of C_m,

$$C_m(s) = o^*(\lambda, s) = \bar{o}(s[1..m])\bar{o}(s[m+1..2m]) \ldots \bar{o}(s[(p-1)m+1..pm]).$$

And, using the definition of $\bar{o}$,

$$|C_m(s)| = |o^*(\lambda, s)| = \sum_{j=1}^{p} |\bar{o}(s[(j-1)m+1..jm])| \leq |s|/m + \sum_{y\in\mathcal{A}^n} \mathrm{occ}(y,s)(-\log_{|\mathcal{B}|} f_y).$$

We obtain the following upper bound for the compression ratio for C_m of α^n,

$$\begin{aligned} \rho_{C_m}(\alpha^n) &= \liminf_{k\to\infty} \frac{|C_m(\alpha^n[1..k])|}{\log_{|\mathcal{B}|} |\mathcal{A}^n|\ k} \\ &\leq \lim_{k\to\infty} \frac{|C_m(\alpha^n[1..i_k])|}{\log_{|\mathcal{B}|} |\mathcal{A}^n|\ i_k} \\ &\leq \lim_{k\to\infty} \left(\frac{i_k}{m} + \sum_{y\in\mathcal{A}^n} \mathrm{occ}(y, \alpha^n[1..i_k])(-\log_{|\mathcal{B}|} f_y) \right) \Big/ \left(\log_{|\mathcal{B}|} |\mathcal{A}^n|\ i_k \right) \\ &\leq \frac{1}{m\ \log_{|\mathcal{B}|} |\mathcal{A}^n|} + \sum_{y\in\mathcal{A}^n} f_y(-\log_{|\mathcal{B}|} f_y)/\log_{|\mathcal{B}|} |\mathcal{A}^n|. \end{aligned}$$

Since we assumed there was some $x \in \mathcal{A}^n$ such that $f_x \neq |\mathcal{A}|^{-n}$, by Shannon's work [13] we have

$$\sum_{y\in\mathcal{A}^n} f_y(-\log_{|\mathcal{B}|} f_y)/\log_{|\mathcal{B}|} |\mathcal{A}^n| < 1.$$

Then, for some sufficiently large m, $\rho_{C_m}(\alpha^n)$ is also strictly less than 1. This proves that α^n, as a sequence in $(\mathcal{A}^n)^\omega$, is compressible. By the invariance of the compressibility ratio under powers of the alphabet, shown in Lemma 1, the sequence $\alpha \in \mathcal{A}^\omega$ is also compressible. This concludes the proof of the Characterization Theorem.

4 Proof of Agafonov's Theorem

As a corollary of the Characterization Theorem we obtain Agafonov's Theorem [1]. We regard a finite automaton that selects a subsequence of a given sequence as a finite-state compressor that behaves as the identity function but *only* on selected positions. We call it a *finite-state selector*.

Definition. A *finite-state selector* is a 5-uple $S = \langle \mathcal{A}, \mathcal{Q}, q_0, \delta, \mathcal{Q}_f \rangle$ where $\mathcal{A}$ is the input alphabet, $\mathcal{Q}$ is a finite set of states, $q_0 \in \mathcal{Q}$ is the initial state, $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is the transition function and $\mathcal{Q}_f \subseteq \mathcal{Q}$ is the set of selecting states. To ensure that for an infinite input the output is necessarily infinite, we require that the transition function δ be free of cycles of non-selecting states. The output function $o : \mathcal{Q} \times \mathcal{A} \to \mathcal{A}$ is the identity function restricted to the selecting states: $o(q, x) = x$, if $q \in \mathcal{Q}_f$; and $o(q, x) = \lambda$, otherwise. The automaton processes the input symbols according to the current state q. When a symbol $x \in \mathcal{A}$ is read, the automaton moves to state $\delta(q, x)$ and outputs x if $q \in \mathcal{Q}_f$; otherwise, it outputs nothing. We define the extensions δ^* and o^* as in the definition of finite-state compressors. We write $S(s)$ for $o^*(q_0, s)$, the output of the selector on input the string s.

The next lemma is a known result in the area (see for instance [9]); it asserts that finite-state selectors can not select a sublinear part of the input.

Lemma 2. *Let $\alpha \in \mathcal{A}^\omega$ and let S be a finite-state selector with k states, then $\rho_S(\alpha) \geq 1/k$.*

Proof. Let $S = \langle \mathcal{A}, \mathcal{Q}, q_0, \delta, \mathcal{Q}_f \rangle$ and $(q_n)_{n \in \mathbb{N}}$ be the sequence of states visited when processing α, $q_n = \delta^*(q_0, \alpha[1..n])$. Consider blocks of $k = |\mathcal{Q}|$ consecutive states in the sequence $q_t, q_{t+1}, ..., q_{t+k-1}$. Since there are k states in a block either all states appear in it or there is a cycle of states. Either way, at least one state in the block must be a selecting state; therefore, at least one symbol from each block of k consecutive symbols of α must be selected. □

We now introduce a technical tool that we will use in the proof of Agafonov's theorem: a finite-state compressor with more than one output. We show that the compression ratio for ordinary finite-state compressors and for these new ones coincides.

Definition. A *two-output finite-state compressor* is a 7-tuple $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o_1, o_2 \rangle$ where $\mathcal{A}$ is the input alphabet, $\mathcal{Q}$ is a finite set of states, $q_0 \in \mathcal{Q}$ is the initial state, $\delta : \mathcal{Q} \times \mathcal{A} \to \mathcal{Q}$ is the transition function and $o_i : \mathcal{Q} \times \mathcal{A} \to \mathcal{B}^*$ are the output functions. The automaton processes the input symbols according to the current state q. When a symbol $x \in \mathcal{A}$ is read, the automaton moves to state $\delta(q, x)$ and outputs $o_i(q, x)$ on output tape i. We extend δ and o_i to process strings in the same way as for regular finite-state compressors. We write $C(s)$ for $\langle o_1^*(q_0, s), o_2^*(q_0, s) \rangle$ and

$$|C(s)| = |o_1^*(q_0, s)| + |o_2^*(q_0, s)|.$$

A two-output finite-state compressor is *lossless* if, from two given output strings and a finishing state, there is at most one input that produces from the initial state the two strings and the finishing state. This is equivalent to requiring that the function $f(s) = \langle o_1^*(q_0, s), o_2^*(q_0, s), \delta^*(q_0, s) \rangle$ be injective.

The definition of compression ratio for a two-output finite-state compressor is exactly as the definition for the case of a single output.

Lemma 3. *The finite-state compression ratio of a given infinite sequence is equal to the two-output compression ratio of the same sequence.*

Proof. Any lossless compressor can be emulated by a two-output lossless compressor by not using one of the outputs. Therefore, the two-output compression ratio is clearly less than or equal to the finite-state compression ratio. Let us show that for any lossless two-output compressor $C = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o_1, o_2 \rangle$ and an infinite sequence $\alpha \in \mathcal{A}^\omega$ we can build a lossless finite-state compressor C' such that $\rho_{C'}(\alpha)$ is arbitrarily close to $\rho_C(\alpha)$.

The idea is to interleave both outputs in blocks of m symbols with one extra symbol before each block that identifies which output it came from. Let $b_1, b_2 \in \mathcal{B}$ be different symbols. We will use b_i to mark that a given output block comes from output i. Let $F_m, L_m : \mathcal{B}^* \to \mathcal{B}^*$ be the functions that split the output such that

$$F_m(s) = s[1..|s| - |s| \mod m] \quad \text{and} \quad L_m(s) = [|s| - |s| \mod m + 1..|s|].$$

Clearly, for all s, $F_m(s)L_m(s) = s$, $|F_m(s)| \mod m = 0$, and $|L_m(s)| < m$. Let $F'_m : \mathcal{B}^* \times \mathcal{B} \to \mathcal{B}^*$ be equal to F but appending the symbol b before each block of m symbols. Thus,

$$F'_m(s, b) = \prod_{i=1}^{|F_m(s)|/m} b\ F_m(s)[im + 1..(i + 1)m].$$

Let $C_m = \langle \mathcal{A}, \mathcal{B}, \mathcal{Q} \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}, \langle q_0, \lambda, \lambda \rangle, \delta'_m, o'_m \rangle$ where

$$\begin{aligned} \delta'_m(\langle q, t_1, t_2 \rangle, x) &= \langle \delta(q, x), L_m(t_1 o_1(q, x)), L_m(t_2 o_2(q, x)) \rangle \\ o'_m(\langle q, t_1, t_2 \rangle, x) &= F'_m(t_1 o_1(q, x), b_1) F'_m(t_2\ o_2(q, x), b_2). \end{aligned}$$

Notice that C_m basically mimics the behavior of C and puts in its single output both outputs of C in blocks of m bits, each preceded with an indicator symbol b_i to indicate that the block came from output i.

Consider a fixed m and let us show that C_m is lossless. Let $f_i : \mathcal{B}^* \times (\mathcal{Q} \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}) \to \mathcal{B}^*$ and $g : \mathcal{B}^* \times (\mathcal{Q} \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}) \to \mathcal{Q}$ be functions that, given an output and finishing state of C_m, calculate both outputs and the finishing state, respectively, of C. From the existence of such functions, since C is lossless, it is clear that C_m is also lossless. Let

$$J_{t,i} = \{j : 1 \le j \le |t|/(m + 1) \wedge t[j(m + 1) - m] = b_i\}$$

be the set of positions of blocks in t that come from output i. Then,

$$\begin{aligned} f_i(t, q, u_1, u_2) &= \left(\prod_{j \in J_{t,i}} t[j(m + 1) - m + 1..j(m + 1)] \right) u_i \\ g(t, q, u_1, u_2) &= q \end{aligned}$$

Simply f_i parses t into blocks of $m + 1$ bits, and it appends to its output only the final m bits of each block starting with marker b_i. Finally, notice that

$$\rho_{C_m}(\alpha) = \liminf_{n \to \infty} \frac{C_m(\alpha[1..n])}{n} \le \liminf_{n \to \infty} \frac{m + 1}{m} \frac{|o_1^*(\alpha[1..n])| + |o_2^*(\alpha[1..n])|}{n} \le \left(1 + \frac{1}{m}\right) \rho_C(\alpha).$$

Letting $C' = C_m$ for sufficiently large m we can make the compression ratio for C' be arbitrarily close to the compression ratio for C. □

We are ready to prove Agafonov's Theorem generalized to arbitrary alphabets.

Agafonov's Theorem. *A sequence $\alpha \in \mathcal{A}^\omega$ is normal to alphabet $\mathcal{A}$ if, and only if, every finite-state selector on input α outputs a sequence normal to alphabet $\mathcal{A}$.*

Proof. In this proof we use finite-state compressors whose output alphabet is the same as the input alphabet, to match the input/output behavior of finite-state selectors.

Assume α is normal to alphabet $\mathcal{A}$ and, towards a contradiction, suppose $S = \langle \mathcal{A}, \mathcal{Q}_S, q_{0S}, \delta_S, \mathcal{Q}_f \rangle$ is a selector such that $S(\alpha)$ is not normal to alphabet $\mathcal{A}$. By the Characterization Theorem, there is a lossless $C = \langle \mathcal{A}, \mathcal{A}, \mathcal{Q}_C, q_{0C}, \delta_C, o_C \rangle$ with output alphabet $\mathcal{A}$ and a positive ε_C such that

$$\rho_C(S(\alpha)) = \liminf_{n \to \infty} |C(S(\alpha)[1..n])|/n = 1 - \varepsilon_C.$$

We define a two-output compressor C' that runs C on the subsequence selected by S, and acts as the identity function on the rest of the input sequence. Let $C' = \langle \mathcal{A}, \mathcal{A}, \mathcal{Q}_C \times \mathcal{Q}_S, \langle q_{0C}, q_{0S} \rangle, \delta, o_1, o_2 \rangle$, where

$$\begin{aligned}
\delta(\langle q_C, q_S \rangle, x) &= \begin{cases} \langle q_C, \delta_S(q_S, x) \rangle & \text{, if } q_S \notin \mathcal{Q}_f \\ \langle \delta_C(q_C, x), \delta_S(q_S, x) \rangle & \text{, if } q_S \in \mathcal{Q}_f. \end{cases} \\
o_1(\langle q_C, q_S \rangle, x) &= \begin{cases} \lambda & \text{, if } q_S \notin \mathcal{Q}_f \\ o_C(q_C, x) & \text{, if } q_S \in \mathcal{Q}_f. \end{cases} \\
o_2(\langle q_C, q_S \rangle, x) &= \begin{cases} x & \text{, if } q_S \notin \mathcal{Q}_f \\ \lambda & \text{, if } q_S \in \mathcal{Q}_f. \end{cases}
\end{aligned}$$

By construction, C' is lossless, because it reproduces the input in one case, and it applies a lossless compressor in the other case.

Let $\bar{S}$ be exactly as the selector $S = \langle \mathcal{A}, \mathcal{Q}_S, q_{0S}, \delta_S, \mathcal{Q}_f \rangle$ but complementing the selecting states. This is $\bar{S} = \langle \mathcal{A}, \mathcal{Q}_S, q_{0S}, \delta_S, \mathcal{Q} \setminus \mathcal{Q}_f \rangle$. Observe that for $s \in \mathcal{A}^*$, $C'(s) = \langle C(S(s)), \bar{S}(s) \rangle$, so

$$|C'(s)| = |C(S(s))| + |\bar{S}(s)| = |C(S(s))| + |s| - |S(s)|.$$

Then,

$$\rho_{C'}(s) \leq \rho_C(S(s))\rho_S(s) + (1 - \rho_S(s))$$

and for $\alpha \in \mathcal{A}^\omega$,

$$\rho_{C'}(\alpha) \leq \rho_C(S(\alpha))\rho_S(\alpha) + (1 - \rho_S(\alpha)).$$

By Lemma 2, $\rho_S(\alpha) \geq \varepsilon_S$ for a positive $\varepsilon_S = 1/|\mathcal{Q}_S|$. By definition of C, $\rho_C(S(\alpha)) = (1 - \varepsilon_C)$ for some positive ε_C. Since both constants are positive,

$$\rho_{C'}(\alpha) \leq (1 - \varepsilon_C)\varepsilon_S + (1 - \varepsilon_S) < 1.$$

Thus, C' compresses α and, by Lemma 3, α is compressible by ordinary lossless finite-state compressors. This contradicts the Characterization Theorem because we assumed that α is normal to alphabet $\mathcal{A}$.

The other direction of the theorem is ensured by the finite-state selector that selects all the symbols of the input sequence. □

Acknowledgements. We are thankful to Elvira Mayordomo and Ludwig Staiger for providing us with a detailed account of the history of the Characterization Theorem and to an anonymous referee for suggesting many improvements in the presentation of the results.

The authors are members of the Laboratoire International Associé INFINIS, Universidad de Buenos Aires – Université Paris Diderot. This research was partially done whilst the first author was a visiting fellow at the Isaac Newton Institute for Mathematical Sciences in the programme 'Semantics & Syntax'.

References

[1] V. N. Agafonov. Normal sequences and finite automata. *Soviet Mathematics Doklady*, 9:324–325, 1968.

[2] É. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909.

[3] C. Bourke, J. Hitchcock, and N. Vinodch. Entropy rates and finite-state dimension. *Theoretical Computer Science*, 349:392–406, 2005.

[4] A. Broglio and P. Liardet. Predictions with automata. symbolic dynamics and its applications. *Contemporary Mathematics*, (135):111–124, 1992. Also in Proceedings AMS Conference in honor of R. L. Adler. New Haven CT - USA 1991.

[5] Y. Bugeaud. *Distribution Modulo One and Diophantine Approximation*. Series: Cambridge Tracts in Mathematics 193. Cambridge University Press, 2012.

[6] J. Dai, J. Lathrop, J. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310:1–33, 2004.

[7] D. Huffman. A method for the construction of minimum-redundancy codes. In *Institute of Radio Engineers*, pages 1098–1102, 1952.

[8] D. Huffman. Canonical forms for information-lossless finite-state logical machines. *Transactions on Information Theory*, 5(5):41–59, 1959.

[9] R. Lindner and L. Staiger. *Algebraische Codierungstheorie – Theorie der sequentiellen Codierungen*. Akademie-Verlag, Berlin, 1977.

[10] W. Merkle and J. Reimann. Selection functions that do not preserve normality. *Theory of Computing Systems*, 39(5):685–697, 2006.

[11] M. G. O'Connor. An unpredictability approach to finite-state randomness. *Journal of Computer and System Sciences*, 37(3):324–336, 1988.

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

[13] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 1948.