Normal Numbers and Finite Automata

Verónica Becher  Pablo Ariel Heiber
vbecher@dc.uba.ar  pheiber@dc.uba.ar
Departamento de Computación, Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires & CONICET, Argentina

January 30, 2013

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.

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 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 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 \( \alpha \) 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
(languages recognized by one-turn pushdown automata, namely, the automata with limited oper-
ations 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 \( \mathcal{A} \), \( \mathcal{A}^\leq n = \bigcup_{i=0}^{n-1} \mathcal{A}^i \) 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 \( |\mathcal{A}^n| = |\mathcal{A}|^n \). \( \lambda \) 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\(^1\); 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 \( \alpha 
\[
\forall x \in \mathcal{A}, \quad \lim_{k \to \infty} \frac{\text{occ}(x, \alpha[1..k])}{k} = \frac{1}{|\mathcal{A}|},
\]
where \( \text{occ}(x, s) = |\{i : x = s[i]\}| \) 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 
\(^1\)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 \( \{0,1,\ldots,b-1\} \) 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 \( r, br, b^2r \ldots \) is 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].
Thus, normality to a given alphabet implies normality to any power of that alphabet. Also notice that 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 $C = (A, B, Q, q_0, \delta, o)$ where $A$ is the input alphabet, $B$ is the output alphabet, $Q$ is a finite set of states, $q_0 \in Q$ is the initial state, $\delta : Q \times A \rightarrow Q$ is the transition function and $o : Q \times A \rightarrow B^*$ is the output function. The automaton processes the input symbols according to the current state and outputs $o(q, x)$. When a symbol $x \in A$ is read, the automaton moves to state $\delta(q, x)$ and outputs $o(q, x)$.

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 $f(s) = (o^*(q_0, s), \delta^*(q_0, s))$ be injective.

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

$$\rho_C(s) = \frac{|C(s)|}{|s| \log_B |A|}.$$

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

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

The finite-state compression ratio of a given sequence $\alpha$ 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 $s$.

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

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

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

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

To go from $A$ to $A^n$, we combine the processing of $n$ consecutive symbols into one transition. Let $C = (A, B, Q, q_0, \delta, o)$ be a finite-state compressor. For a given $n$, we define $C^n =$
\( (\mathcal{A}^n, \mathcal{B}, \mathcal{Q}, q_0, \delta^*, o^*) \) by restricting the domain of \( \delta^* \) 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 \( \mathcal{A}^0 \) 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 \( C^n = (\mathcal{A}^n, \mathcal{B}, \mathcal{Q}, q_0, \delta, o) \), we define \( C = (\mathcal{A}, \mathcal{B}, \mathcal{Q} \times \mathcal{A}^{<n}, (q_0, \lambda), \delta', o') \), where

\[
\delta'(q, s, x) = \begin{cases} (q, sx), & \text{if } |s| < n - 1, \\ (\delta(q, sx), \lambda), & \text{if } |s| = n - 1. \end{cases}
\]

\[
o'(q, s, x) = \begin{cases} \lambda, & \text{if } |s| < n - 1, \\ o(q, sx), & \text{if } |s| = n - 1. \end{cases}
\]

The fact that \((\delta')^*(q, \lambda) = (\delta(q, s), \lambda) \) and \((o')^*((q, \lambda), s) = o^*(q, s) \) 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 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 successively, one does not recover the same automaton (because one of the transformations changes the set of states and the other does not).

### 3 Proof of the Characterization Theorem

#### 3.1 Normal Implies Incompressible

Assume \( \alpha \in \mathcal{A}^* \) is normal to alphabet \( \mathcal{A} \). Let \( C = (\mathcal{A}, \mathcal{B}, \mathcal{Q}, q_0, \delta, o) \) 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 \( \alpha \) is strictly larger than \((1 - \varepsilon)^3\).

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:

\[
S_n = \{ s \in \mathcal{A}^n : a_s > (1 - \varepsilon)n \log_{|\mathcal{B}|} |\mathcal{A}| \}.
\]

Given a string \( s \) in the complement set \( \mathcal{A}^n \setminus S_n \) 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 \( f(s) = (q_s, \delta^*(q_s, s), o^*(q_s, s)) \), for an appropriate \( q_s \) defines an injective function

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

Thus, we can bound the cardinalities of \( \mathcal{A}^n \setminus S_n \) and \( S_n \) as follows:

\[
|\mathcal{A}^n \setminus S_n| < |\mathcal{Q}|^2 |\mathcal{B}|^{(1 - \varepsilon)n \log_{|\mathcal{B}|} |\mathcal{A}| + 1} = |\mathcal{B}| |\mathcal{Q}|^2 |\mathcal{A}|^{(1 - \varepsilon)n},
\]

\[
|S_n| > |\mathcal{A}^n| - |\mathcal{B}| |\mathcal{Q}|^2 |\mathcal{A}|^{(1 - \varepsilon)n} = |\mathcal{A}^n| (1 - |\mathcal{B}| |\mathcal{Q}|^2 |\mathcal{A}|^{-\varepsilon n}).
\]

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 \(|S_n| > |\mathcal{A}^n| (1 - \varepsilon)\).
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 \( \alpha^n \in (\mathcal{A}^n)^n \) be the sequence \( \alpha \) seen in alphabet \( \mathcal{A}^n \), that is, \( \alpha^n[i] = \alpha[(i-1)n + 1..in] \). By the definition of normality, since \( \alpha \) is normal to alphabet \( \mathcal{A} \), \( \alpha^n \) is simply normal to alphabet \( \mathcal{A}^n \). Then, let \( k_0 \) be such that
\[
\forall k > k_0 \, \forall x \in \mathcal{A}^n \, \frac{\text{occ}(x, \alpha^n[1..k])}{k} > |\mathcal{A}^n|^{-1} (1 - \varepsilon) .
\]
Observe that \( |\mathcal{A}^n|^{-1} (1 - \varepsilon) = |\mathcal{A}|^{-n} (1 - \varepsilon) \). 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 \( s \in S_n \), 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 \( \alpha^n[1..k] \):
\[
|C^n(\alpha^n[1..k])| \geq \sum_{x \in \mathcal{A}^n} \text{occ}(x, \alpha^n[1..k]) \cdot a_x
\]
\[
> \sum_{x \in S_n} k \cdot |\mathcal{A}|^{-n} (1 - \varepsilon) \cdot a_x
\]
\[
> \sum_{x \in S_n} k \cdot |\mathcal{A}|^{-n} (1 - \varepsilon)(n \log |\mathcal{B}| |\mathcal{A}|) = |S_n| k |\mathcal{A}|^{-n} (1 - \varepsilon)^2 \log |\mathcal{B}| |\mathcal{A}|^n
\]
\[
> (1 - \varepsilon)^3 k \log |\mathcal{B}| |\mathcal{A}|^n.
\]
Therefore, the compression ratio for \( C^n \) of \( \alpha^n \), \( \rho_{C^n}(\alpha^n) = \liminf_{k \to \infty} \frac{|C^n(\alpha^n[1..k])|}{k \log |\mathcal{B}| |\mathcal{A}|^n} \) 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 \( \alpha \in \mathcal{A}^n \) is not normal to alphabet \( \mathcal{A} \). We will show that \( \alpha \) is compressible (regardless of the choice of an output alphabet \( \mathcal{B} \)). Since \( \alpha \) is not normal to alphabet \( \mathcal{A} \), there is some \( n \) such that \( \alpha \) is not simply normal to alphabet \( \mathcal{A}^n \). Fix such a block length \( n \). As before, let \( \alpha^n \) 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
\[
\lim_{k \to \infty} \frac{\text{occ}(x, \alpha^n[1..k])}{k} \neq \frac{1}{|\mathcal{A}|}.
\]

Either this limit does not exist, or it is different from \(|\mathcal{A}|^{-n}\). Thus, it is impossible that both,
\[
\liminf_{k \to \infty} \frac{\text{occ}(x, \alpha^n[1..k])}{k} \quad \text{and} \quad \limsup_{k \to \infty} \frac{\text{occ}(x, \alpha^n[1..k])}{k},
\]
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 \( y \in \mathcal{A}^n \), the limiting frequency of \( y \) at positions \((i_k)_{k \in \mathbb{N}}\),
\[
f_y = \lim_{k \to \infty} \frac{\text{occ}(x, \alpha^n[1..i_k])}{i_k}
\]
is defined and \( f_x \neq |\mathcal{A}^n|^{-1} \). 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 \((i_k^{(1)})_{k \in \mathbb{N}}\) be an increasing sequence of positions such that \( f_{y_1} \) is defined and different from \(|\mathcal{A}^n|^{-1}\). This exists because \( y_1 = x \) and we already argued that the limit for \( x \) is not \(|\mathcal{A}^n|^{-1}\) over all subsequences. And for each \( j, 2 \leq j \leq |\mathcal{A}^n| \), let \((i_k^{(j)})_{k \in \mathbb{N}}\) be a subsequence of \((i_k^{(j-1)})_{k \in \mathbb{N}}\) such that the limit \( f_{y_j} \) is defined when considered over positions \((i_k^{(j)})\). Since frequencies are bounded between 0 and 1,
such a subsequence necessarily exists. Observe that the sequence \( (\frac{q}{i_k(A^n)})_{k \in \mathbb{N}} \) verifies that for all \( y \in A^n \), \( f_y \) is defined. By letting \( i_k = \frac{q}{i_k(A^n)} \), for each \( k \in \mathbb{N} \), we obtain the desired sequence.

We now prove that \( \alpha^n \) is compressible. We shall bound the compression ratio of \( \alpha^n \) 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 \( C_m = (A^n, B, (A^n)^<m, \lambda, \delta, o) \) be such that for each \( q \in (A^n)^<m \) and \( z \in A^n \),

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

\[
o(q, z) = \begin{cases} 
\lambda, & \text{if } |q| < m - 1, \\
o(qz), & \text{if } |q| = m - 1,
\end{cases}
\]

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

\[
|o(z_1z_2\ldots z_m)| = \left\lceil \sum_{i=1}^m -\log_{|B|} f_{z_i} \right\rceil.
\]

Since

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

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 \( s \in (A^n)^* \). Fix \( p \) to be the largest integer such that \( pm \leq |s| \). By definition of \( C_m \),

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

And, using the definition of \( o \),

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

We obtain the following upper bound for the compression ratio for \( C_m \) of \( \alpha^n \),

\[
\rho_{C_m}(\alpha^n) = \lim_{k \to \infty} \inf_{(A^n)^{<k}} \frac{|C_m(\alpha^n[1..k])|}{\log_{|B|} |A^n| k} \\
\leq \lim_{k \to \infty} \frac{|C_m(\alpha^n[1..i_k])|}{\log_{|B|} |A^n| i_k} \\
\leq \lim_{k \to \infty} \left( \frac{i_k}{m} + \sum_{y \in A^n} \text{occ}(y, \alpha^n[1..i_k])(-\log_{|B|} f_y) \right) / \left( \log_{|B|} |A^n| i_k \right) \\
\leq \frac{1}{m} \log_{|B|} |A^n| + \sum_{y \in A^n} f_y(-\log_{|B|} f_y) / \log_{|B|} |A^n|.
\]

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

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

Then, for some sufficiently large \( m \), \( \rho_{C_m}(\alpha^n) \) is also strictly less than 1. This proves that \( \alpha^n \), as a sequence in \((A^n)^*\), is compressible. By the invariance of the compressibility ratio under powers of the alphabet, shown in Lemma 1, the sequence \( \alpha \in A^w \) 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-tuple \(S = ⟨A, Q, q_0, δ, Q_f⟩\) where \(A\) is the input alphabet, \(Q\) is a finite set of states, \(q_0 \in Q\) is the initial state, \(δ : Q \times A → Q\) is the transition function and \(Q_f \subseteq Q\) 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 \(o : Q \times A → A\) is the identity function restricted to the selecting states: \(o(q,x) = x\), if \(q ∈ Q_f\); and \(o(q,x) = λ\), otherwise. The automaton processes the input symbols according to the current state \(q\). When a symbol \(x ∈ A\) is read, the automaton moves to state \(δ(q,x)\) and outputs \(x\) if \(q ∈ Q_f\); 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, see for instance [9]; it asserts that finite-state selectors can not select a sublinear fraction of the input.

Lemma 2. Let \(α ∈ A^∗\) and let \(S\) be a finite-state selector with \(k\) states, then \(ρ_S(α) ≥ 1/k\).

Proof. Let \(S = ⟨A, Q, q_0, δ, Q_f⟩\) and \((q_n)_{n∈N}\) be the sequence of states visited when processing \(α\), \(q_n = δ^∗(q_0,α[1..n])\). Consider blocks of \(k\) 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 \(C = ⟨A, B, Q, q_0, δ, o_1, o_2⟩\) where \(A\) is the input alphabet, \(Q\) is a finite set of states, \(q_0 ∈ Q\) is the initial state, \(δ : Q \times A → Q\) is the transition function and \(o_1 : Q \times A → B^∗\) and \(o_2 : Q \times A → B^∗\) are the output functions. The automaton processes the input symbols according to the current state \(q\). When a symbol \(x ∈ A\) is read, the automaton moves to state \(δ(q,x)\) 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 \(⟨o_1^∗(q_0,s), o_2^∗(q_0,s)⟩\) 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 \(f(s) = ⟨o_1^∗(q_0,s), o_2^∗(q_0,s), δ^∗(q_0,s)⟩\) 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 \(C = ⟨A, B, Q, q_0, δ, o_1, o_2⟩\) and an infinite sequence \(α ∈ A^∗\) we can build a lossless finite-state compressor \(C′\) such that \(ρ_{C′}(α)\) is arbitrarily close to \(ρ_C(α)\).
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 \( C_m = \langle A, B, Q \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}, (q_0, \lambda, \lambda), \delta'_m, \alpha'_m \rangle \) where
\[
\delta'_m((q,t_1,t_2), x) = \langle \delta(q, x), L_m(t_1, q, x)), L_m(t_2, q, x) \rangle
\]
\[
\alpha'_m((q,t_1,t_2), x) = F'_m(t_1, q, x), b_1 F'_m(t_2, q, x), b_2).
\]
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 \( f_i : \mathcal{B}^* \times (Q \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}) \to \mathcal{B}^* \) and \( g : \mathcal{B}^* \times (Q \times \mathcal{B}^{<m} \times \mathcal{B}^{<m}) \to \mathcal{Q} \) 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
\[
J_{i,i} = \{j : 1 \leq j \leq |t|/(m+1) \land t([j(m+1) - m] = b_i)\}
\]
be the set of positions of blocks in \( t \) that come from output \( i \). Then,
\[
f_i(t, q, u_1, u_2) = \left( \prod_{j \in J_{i,i}} t[j(m+1) - m + j(m+1)] \right) u_i
\]
\[
g(t, q, u_1, u_2) = q.
\]
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
\[
\rho_{C_m}(\alpha) = \liminf_{n \to \infty} \frac{C_m(\alpha[1..n])}{n} \leq \liminf_{n \to \infty} \frac{m + 1}{m} \frac{|\alpha'_1(\alpha[1..n])| + |\alpha'_2(\alpha[1..n])|}{n} \leq \left( 1 + \frac{1}{m} \right) \rho_{C}(\alpha).
\]
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 \( \alpha \in \mathcal{A}^* \) is normal to alphabet \( \mathcal{A} \) if, and only if, every finite-state selector on input \( \alpha \) 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 \( \alpha \) is normal to alphabet \( \mathcal{A} \) and, towards a contradiction, suppose \( S = \langle A, \mathcal{Q}S, q_0S, \deltaS, \mathcal{Q}_{f} \rangle \) is a selector such that \( S(\alpha) \) is not normal to alphabet \( \mathcal{A} \). By the Characterization Theorem, there is a lossless \( C = \langle A, \mathcal{A}, \mathcal{Q}C, q_0C, \deltaC, \mathcal{Q}_{C} \rangle \) with output alphabet \( \mathcal{A} \) and a positive \( \varepsilon_C \) such that
\[
\rho_{C}(S(\alpha)) = \liminf_{n \to \infty} \frac{|C(S(\alpha)[1..n])|}{n} = 1 - \varepsilon_C.
\]
We define a two-output compressor $C'$ that runs $C$ on the subsequence selected by $S$, and acts as the identity on the rest of the input sequence. Let $C' = (\mathcal{A}, \mathcal{A}, \mathcal{Q}_S \times \mathcal{Q}_S, (q_{0C}, q_{0S}), \delta, o_1, o_2)$, where

$$
\delta((q_C, q_S), x) = \begin{cases} 
(q_C, \delta_S(q_S, x)) & \text{if } q_S \notin \mathcal{Q}_f \\
(\delta_C(q_C, x), \delta_S(q_S, x)) & \text{if } q_S \in \mathcal{Q}_f.
\end{cases}
$$

$$
o_1((q_C, q_S), x) = \begin{cases} 
\lambda & \text{if } q_S \notin \mathcal{Q}_f \\
o_C(q_C, x) & \text{if } q_S \in \mathcal{Q}_f.
\end{cases}
$$

$$
o_2((q_C, q_S), x) = \begin{cases} 
x & \text{if } q_S \notin \mathcal{Q}_f \\
\lambda & \text{if } q_S \in \mathcal{Q}_f.
\end{cases}
$$

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

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

$$|C'(s)| = |C(S(s))| + |S(s)| = |C(S(s))| + |s| - |S(s)|.$$

Let $i_n$ be an increasing sequence of positions such that

$$\rho_{C'}(\alpha) = \lim_{n \to \infty} \frac{|C'(\alpha[1..i_n])|}{i_n} = \lim_{n \to \infty} \frac{|C(S(\alpha[1..i_n]))| + n - |S(\alpha[1..i_n])|}{n}$$

$$\leq \lim_{n \to \infty} \frac{|C(S(\alpha[1..i_n]))|}{i_n} + 1 - \frac{|S(\alpha[1..i_n])|}{i_n}$$

$$\leq \lim_{n \to \infty} \frac{|C(S(\alpha[1..i_n]))|}{|S(\alpha[1..i_n])|} \rho_S(\alpha) + 1 - \rho_S(\alpha)$$

$$\leq \rho_C(S(\alpha)) \rho_S(\alpha) + 1 - \rho_S(\alpha).$$

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 $\varepsilon_C$. Since both constants are positive, $\rho_{C'}(\alpha) \leq (1 - \varepsilon_C)\varepsilon_S + (1 - \varepsilon_S) < 1$.

Thus, $C'$ compresses $\alpha$ and, by Lemma 3, $\alpha$ is compressible by ordinary lossless finite-state compressors. This contradicts the Characterization Theorem because we assumed that $\alpha$ 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**


