

A Symbolic Dynamical System Approach to Lossy Source Coding with Feedforward

Ofer Shayevitz

University of California, San Diego

La Jolla, CA 92093, USA

ofersha@ucsd.edu

Abstract

It is known that modeling an information source via a symbolic dynamical system evolving over the unit interval, leads to a natural lossless compression scheme attaining the entropy rate of the source, under general conditions. We extend this notion to the lossy compression regime assuming a feedforward link is available, by modeling a source via a two-dimensional symbolic dynamical system where one component corresponds to the compressed signal, and the other essentially corresponds to the feedforward signal. For memoryless sources and an arbitrary bounded distortion measure, we show this approach leads to a family of simple deterministic compression schemes that attain the rate-distortion function of the source. The construction is dual to a recent optimal scheme for channel coding with feedback.

I Introduction

Lossless compression of a discrete information source to its entropy rate \mathcal{H} is a well studied topic. A possibly lesser known approach to this problem is one based on symbolic dynamical systems, where the information generating mechanism is modeled by a randomly initialized iterative mapping of the unit interval to itself, and the emitted source sequence is a quantized observation of that process. For well behaved mappings the source sequence constitutes an *expansion* of the initial point, i.e., corresponds to a unique such point. Furthermore, the prefixes of this expansion describe the initial point with (exponentially) increasing resolution, and the unit interval can be uniformly partitioned into $\approx 2^{n\mathcal{H}}$ subintervals so that with high probability, the subinterval containing the initial point will have all its points admitting the same length- n expansion. This leads to a conceptually simple and optimal compression scheme: A finite source sequence is mapped to a representing subinterval by computing the corresponding reverse trajectory of the dynamical system, and is reconstructed by following the trajectory of an arbitrary point in that subinterval¹. A comprehensive study of the symbolic dynamics framework for information sources can be found in [1]. Some of the ideas can be traced back to Rényi, see [2] and references therein.

¹This has a flavor similar to arithmetic coding and (using variable-length coding) essentially coincides with it in some cases, see Example 1.

In this paper, we extend the concept above to the lossy source coding regime, under the assumption that a noiseless feedforward link is available. This setting is described as follows: An encoder observes a stochastic source sequence Y^n over some product alphabet \mathcal{Y}^n , and maps it to a rate R index set $E = \{1, 2, \dots, 2^{nR}\}$ using some encoding function $e : \mathcal{Y}^n \mapsto E$. The index is sent to the decoder. At time k , the decoder knows the sequence Y^{k-1} via the feedforward link, and generates an approximation of Y_k using a decoding function $f_k : E \times \mathcal{Y}^{k-1} \mapsto \mathcal{X}$, where \mathcal{X} is the reconstruction alphabet. The quality of the approximation is measured w.r.t. a distortion measure $d : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}^+$, by evaluating the time-averaged expected distortion:

$$D = n^{-1} \sum_{k=1}^n \mathbb{E} d(f_k(e(Y^n), Y^{k-1}), Y_k)$$

The *rate-distortion function* of the source is the infimum of all rates R for which there exist encoding and decoding functions achieving a distortion at most D , for any n large enough. It is denoted $R_{\text{ff}}(D)$ under the feedforward assumption, and $R(D)$ where feedforward is absent (i.e., when restricting $f_k(e, y^{k-1}) = f_k(e)$).

This model has been initially motivated and studied in the context of competitive prediction [3], where it was shown that feedforward does not decrease the rate-distortion function for a large family of sources (in particular, memoryless). An in-depth analysis of the rate-distortion function with feedforward appears in [4]. A simple scheme inspired by a successive error compression feedback coding technique and achieving the rate-distortion function for discrete memoryless sources, was suggested in [5]. Another optimal protocol building on the Schalkwijk-Kailath scheme for channel coding with feedback over the AWGN, was suggested for the white Gaussian source [6]. In this paper, we suggest an alternative approach based in symbolic dynamics and motivated by a recent optimal feedback transmission scheme, termed *posterior matching* [7][8][9]. The suggested approach yields a conceptually simple compression protocol, which is shown to achieve the rate-distortion function for discrete memoryless sources with a bounded distortion measure.

II Preliminaries

Random variables (r.v.'s) are denoted by upper-case letters, their realizations by corresponding lower-case letters. A r.v. X (either real or discrete) is associated with a probability distribution $P_X(\cdot)$ (over \mathbb{R} , or over a discrete alphabet $\mathcal{X} \subseteq \mathbb{N}$) and we write $X \sim P_X$. The *cumulative distribution function* (c.d.f.) of X is denoted by F_X . We write $\mathbb{E}(\cdot)$ for expectation and $\mathbb{P}(\cdot)$ for the probability of an event within the parentheses. $H(X)$ is the entropy of a discrete r.v. X , $h(X)$ is the differential entropy of a continuous r.v. X , and $I(X; Y)$ is the mutual information between a pair of r.v. X, Y . We use $|\Delta|$ for the length of an interval $\Delta \subseteq \mathbb{R}$, \log for \log_2 , \circ for function composition, \overline{A} for the closure of the set A , $\mathbb{1}_A(\cdot)$ for the indicator function over the set A , \mathcal{I} for the open unit interval $(0,1)$, and $\mathcal{I}_2 \stackrel{\text{def}}{=} \mathcal{I} \times \mathcal{I}$ for the open unit square. An *open partition* of a set A (in what follows, \mathcal{I} or \mathcal{I}_2) is a family of disjoint open subsets $\{A_i\}$ of A , such that $\overline{\cup A_i} = \overline{A}$. A sequence x^n over a finite alphabet is said to be (strongly) ε -typical w.r.t. P_X , if the (zero order) empirical distribution of symbols in x^n is ε -close to the distribution P_X in the supremum norm. The set of all such length n sequences is denoted $\mathcal{T}_{n,\varepsilon}(P_X)$.

We now turn to define a (two-dimensional) *dynamical source*, generalizing the definition in [1]. Note that in the sequel, we discuss in detail a significantly more restrictive

family of dynamical sources. We provide the rather abstract definition below both for future reference, and as we believe it is more instructive.

A dynamical source \mathcal{S} has the following components:

- A triplet of alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathbb{N}$.
- Two open partitions of \mathcal{I} into open intervals $\{\Pi_i^0\}_{i \in \mathcal{X}}$ and $\{\Pi_j^1\}_{j \in \mathcal{Z}}$, and the corresponding product partition $\Pi_{ij} = \Pi_i^0 \times \Pi_j^1$ of \mathcal{I}_2 . Without loss of generality we assume that the intervals are arranged from left to right (or vice versa) according to the natural alphabet order.
- Two functions $\sigma_0 : \mathcal{I} \mapsto \mathcal{X}$, $\sigma_1 : \mathcal{I} \mapsto \mathcal{Z}$ that are equal to i, j over Π_i^0, Π_j^1 respectively.
- A function $\xi : \mathcal{X} \times \mathcal{Z} \mapsto \mathcal{Y}$, and its corresponding extension to $\zeta : \mathcal{I}_2 \mapsto \mathcal{Y}$ that is constant and equal to $\xi(i, j)$ when restricted to Π_{ij} .
- A mapping $T : \mathcal{I}_2 \mapsto \mathcal{I}_2$ of the form

$$T(\theta, \phi) = (T_0(\theta, \zeta(\theta, \phi)), T_1(\theta, \phi)) \quad (1)$$

such that T restricted to each Π_{ij} is a continuous bijection, and $\{T(\Pi_{ij})\}_{(i,j) \in \xi^{-1}(k)}$ is an open partition of \mathcal{I}_2 for each $k \in \mathcal{Y}$.

Setting $(\theta_1, \phi_1) \in \mathcal{I}_2$ as an *initial state*, the source \mathcal{S} is associated with the following sequences, all of which are deterministic functions of the initial state:

- The *state sequence* $(\theta^\infty, \phi^\infty)$ over \mathcal{I}_2 , recursively defined by $(\theta_n, \phi_n) = T(\theta_{n-1}, \phi_{n-1})$.
- The *source sequence* y^∞ over the alphabet \mathcal{Y} , defined by $y_n = \zeta(\theta_n, \phi_n)$
- The *component sequences* x^∞, z^∞ over the alphabets \mathcal{X}, \mathcal{Z} respectively, defined by

$$x_n = \sigma_0(\theta_n), \quad z_n = \sigma_1(\phi_n)$$

Furthermore, any finite source sequence y^n corresponds to a *fundamental set* $u_n(y^n) \subseteq \mathcal{I}_2$, defined to be the set of all initial states $(\theta_1, \phi_1) \in \mathcal{I}_2$ that result in the source sequence y^n .

Following [1] again, a *probabilistic dynamical source* is a pair (\mathcal{S}, P) where \mathcal{S} is a dynamical source, and P is a probability measure equivalent to the Lebesgue measure over \mathcal{I}_2 . Setting $(\Theta_1, \Phi_1) \sim P$ as the initial state, the source (\mathcal{S}, P) is naturally associated with the *stochastic sequences* $\Theta^\infty, \Phi^\infty, Y^\infty, X^\infty, Z^\infty$, all of which are deterministic functions of the initial state.

III Lossless Coding

Let (\mathcal{S}, P) be a probabilistic dynamical source with $|\mathcal{Z}| = 1$, i.e., one dimensional, and we can assume $\mathcal{X} = \mathcal{Y}$. In this case the fundamental sets are simply intervals in \mathcal{I} (in this section we disregard the redundant dimension). Under some further contraction conditions, an asymptotic equipartition property was shown to hold [1], namely $n^{-1} \log |P(u_n(Y^n))|^{-1}$ tends in probability to the entropy rate $\mathcal{H}(Y^\infty)$ of the source

sequence. This immediately leads to an optimal compression protocol: The unit interval is uniformly partitioned into $\approx 2^{n(\mathcal{H}+\delta)}$ representative intervals. The trajectory of the dynamical source is reversed using Y^n , namely recovering the fundamental interval $T_0^{-1}(\cdot, Y_1) \circ \dots \circ T_0^{-1}(\cdot, Y_{n-1}) \circ T_0^{-1}(\mathcal{I}, Y_n)$. The index of a representative contained in the fundamental interval is used to describe the source sequence². To reconstruct Y^n , the dynamical source is initialized with any point inside the representative interval.

Example 1 (Memoryless Sources). To generate a memoryless source over the alphabet \mathcal{Y} , we set $P = \text{Uniform}(\mathcal{I})$, $\xi(i, j) = \xi(i) = i$, and $T_0(\theta, i)$ to be affine and map Π_i^0 to \mathcal{I} . This results in a source sequence that is i.i.d.- P_Y , where $P_Y(i) = |\Pi_i^0|$. If $T_0(\theta, i)$ are all monotonically increasing, the fundamental intervals are precisely those generated by the simple arithmetic coding protocol for the source, and coding them (the typical ones) as described (or alternatively, using a variable-rate code to obtain zero error) results in lossless compression with a rate approaching $H(Y)$. Note that in particular for $Y \sim \text{Uniform}(\mathcal{Y})$, the source sequence is simply the $|\mathcal{Y}|$ -base expansion of the initial state point.

Example 2 (The Continued Fraction Source). The continued fraction expansion of a number in \mathcal{I} can be generated by a dynamical source [1]. In this case we have $\mathcal{Y} = \mathbb{N}$, the open partition is $\Pi_i^0 = (1/(i+1), 1/i)$, $\xi(i, j) = \xi(i) = i$, and $T_0(\theta, i) = \theta^{-1} - i = \theta^{-1} \pmod{1}$. Endowing the source with any probability measure P that is equivalent to the Lebesgue measure over \mathcal{I} , the state process converges to the invariant distribution that admits the density $f_{\text{inv}}(\theta) = \frac{\log e}{1+\theta} \cdot \mathbf{1}_{\mathcal{I}}(\theta)$ [10].

Coding the fundamental intervals as described results in lossless compression with a rate approaching the entropy rate of the continued fraction source, which is given by $\mathcal{H}(Y^\infty) = \frac{\pi^2 \log e}{6}$. It is interesting to note that in this case, a more efficient (yet equivalent) coding mechanism for the fundamental intervals is readily available: Represent a finite source sequence y^n by the unique rational number $\frac{p_n}{q_n}$ it is the continued fraction expansion of. It is well known that for almost all $\theta \in \mathcal{I}$ (w.r.t. the Lebesgue measure), the denominator of the convergents of the continued fraction expansion satisfies $n^{-1} \log q_n \rightarrow \frac{\pi^2 \log e}{12}$ [10], and so (p_n, q_n) can be represented at a rate of twice this number, which is precisely the entropy rate of the continued fraction source.

IV Lossy Coding with Feedforward

IV.1 Motivation

In the lossless setting, a finite source sequence was described by efficiently enumerating (typical) fundamental sets, obtained via a representation of an initial state up to a suitable resolution. In the lossy setting, we wish to provide only partial information regarding the fundamental set. To that end, a two-dimensional dynamical source model was introduced, where the high-level idea is to provide the decoder with a representation of the θ -component of the initial state only. At time k , the decoder knows the sequence Y^{k-1} (via feedforward), and can therefore compute the θ -component Θ_{k-1} that corresponds to the initial state Θ_1 it was given. This is made possible due to the restriction (1) on the structure of T_0 , making its evolution dependent only on the θ -component and the causal knowledge of the source sequence. Had it known the ϕ -component as well,

²If there is no such interval, an arbitrary index is used. This error event is of vanishing probability.

the decoder could have reconstructed X_k, Z_k and hence Y_k . Here, it can only reconstruct X_k , which can serve as an estimate for Y_k .

So, our first task is, for a fixed source sequence distribution, to design a probabilistic dynamical source (\mathcal{S}, P) that is consistent with this distribution, and also makes X^n, Y^n dependent in a prescribed way so that this reconstruction has low distortion. However, there is an even more difficult obstacle. The initial θ -component has to be described with a finite rate, and (loosely speaking) this should be done while making sure that an initial ϕ -component can be selected so that the statistical dependence above is roughly maintained. For memoryless sources, both tasks can be accomplished.

IV.2 Memoryless Sources

Let P_Y be a probability distribution over the alphabet \mathcal{Y} . There are many different probabilistic dynamical sources for which the source sequence is i.i.d.- P_Y . One simple example was given in the previous section, where $|\mathcal{Z}| = 1$ and T is affine on any Π_i^0 , and corresponds to a lossless compression with rate $H(Y)$. However, in two dimensions there is an abundance of distinct probabilistic dynamical sources that admit an i.i.d.- P_Y source sequence.

Consider any channel $P_{X|Y}$ from Y to X over the alphabets $\mathcal{X} \times \mathcal{Y}$, let $P_{XY} = P_Y \times P_{X|Y}$ be the joint distribution and let $P_{Y|X}$ be the corresponding test channel from X to Y . The following Lemma is easily observed [11].

Lemma 1. *There exists an alphabet \mathcal{Z} of size $|\mathcal{Z}| \leq |\mathcal{X}|(|\mathcal{Y}| - 1) + 1$, a function $\xi : \mathcal{X} \times \mathcal{Z} \mapsto \mathcal{Y}$, and a r.v. Z independent of X , such that $(X, \xi(X, Z)) \sim P_{XY}$.*

Now, let us define the following dynamical source \mathcal{S} . The construction is motivated by the *posterior matching scheme*, a capacity achieving feedback transmission scheme for memoryless channels with feedback [7][8][9].

- $|\Pi_i^0| = P_X(i)$ for any $i \in \mathcal{X}$.
- $|\Pi_j^1| = P_Z(j)$ for any $j \in \mathcal{Z}$.
- The function ξ is that of Lemma 1, ζ its natural extension.
- The mapping $T = (T_0, T_1)$ is defined as follows:
 - Let $F_{X|Y}$ be the conditional c.d.f. for $P_{X|Y}$. For any fixed $k \in \mathcal{Y}$, $T_0(\theta, k)$ is a continuous non-decreasing function from \mathcal{I} onto \mathcal{I} , is affine on each Π_i^0 , and is equal to $F_{X|Y}(i|k)$ on the right edge of $\overline{\Pi_i^0}$.
 - $T_1(\theta, \phi) = T_1(\phi)$ is one dimensional, affine on each Π_j^1 and maps it onto \mathcal{I} .

Remark. Note that when $P_{X|Y}$ is noiseless (e.g., $X = Y$) then \mathcal{S} collapses to the one dimensional lossless construction of Example 1.

Lemma 2. *For any $y \in \mathcal{Y}^n$, the fundamental set $u_n(y^n)$ of the dynamical source \mathcal{S} is a finite disjoint union of product rectangles. The projections of these rectangles onto the θ -axis form a set of at most $n(|\mathcal{X}| - 1) + 1$ distinct intervals.*

Proof. The first assertion follows easily from the affinity of T . For $n = 1$, the number of distinct intervals on the θ -axis is exactly $|\mathcal{X}|$. For any fixed $k \in \mathcal{Y}$, $T_0(a, k)$ is quasi-affine over \mathcal{I} as a function of a , with at most $|\mathcal{X}| - 1$ corner points. Hence the number of distinct intervals can increase by at most $|\mathcal{X}| - 1$ at each step. \square

The following Lemma is adapted from [9].

Lemma 3. *Let $P \sim \text{Uniform}(\mathcal{I}_2)$. The probabilistic dynamical source (\mathcal{S}, P) has the following properties:*

- (a) *The sequence Z^∞ is i.i.d.- P_Z , Z_n is statistically independent of X^n .*
- (b) *$(X_n, Y_n) \sim P_{XY}$, X_n is statistically independent of Y^{n-1}*
- (c) *The source sequence Y^∞ is i.i.d.- P_Y , and $Y_n - X_n - X^{n-1}Y^{n-1}Z^{n-1}$ form a Markov chain.*
- (d) *$I(\Theta_1; Y^n) = nI(X; Y)$*

Proof. Assertion (a) is immediate: Z^∞ is a deterministic function of Φ_1 and evolves according to the memoryless dynamical law described in Example 1, hence is i.i.d.- P_Z . Furthermore, X^n is a deterministic function of (Θ_1, Z^{n-1}) which are mutually independent of Z_n . For the other assertions, see [9]. \square

Define

$$\omega_k(\cdot) \stackrel{\text{def}}{=} T_0^{-1}(\cdot, k), \quad \Delta_n^\varepsilon(y^n) \stackrel{\text{def}}{=} \omega_{y_1} \circ \cdots \circ \omega_{y_{n-1}} \circ \omega_{y_n}((\varepsilon, 1 - \varepsilon)) \quad (2)$$

Namely, the interval $\Delta_n^\varepsilon(y^n)$ is obtained by reversing the trajectory of the (edges of the) interval $(\varepsilon, 1 - \varepsilon)$. The following result, also adapted from [9], is central to our derivations.

Theorem 1. *Suppose that P_{XY} is strictly positive over $\mathcal{X} \times \mathcal{Y}$, and that for any fixed $\theta \in \mathcal{I}$, $T_0(\theta, j)$ is not a constant function of j . Then for any $\varepsilon > 0$,*

- (a) *$\lim_{n \rightarrow \infty} \mathbb{P}(|\Delta_n^\varepsilon(Y^n)| > 2^{-nR}) = 0$ for any $R < I(X; Y)$.*
- (b) *$\lim_{n \rightarrow \infty} \mathbb{P}(\Theta_1 \notin \Delta_n^\varepsilon(Y^n)) \leq 2\varepsilon$.*

Loosely speaking, Theorem 1 implies that by observing the source sequence, the initial Θ_1 component of the state sequence can be found up to a resolution of $2^{-nI(X; Y)}$. In a feedback communication setting, this initial value represents a *message* to be sent over the channel $P_{Y|X}$, and this concentration result means that one can reliably transmit roughly $2^{nI(X; Y)}$ such messages and decode them with high reliability, which corresponds to a communication rate of *at most* $I(X; Y)$ bits per channel use. In order to be able to generate X_n (channel input) the encoder needs to know the Y^{n-1} on top of the message Θ_1 , hence the feedback. In the dual lossy source coding with feedforward setting we consider, Θ_1 plays the role of a lossy description of the source sequence, and we will need *at least* $I(X; Y)$ bits per source symbol to represent it with high enough accuracy. In order to be able to generate X_n (lossy reconstruction of Y_n) the decoder needs to know the Y^{n-1} on top of the (quantized representation of the) lossy description Θ_1 , hence the feedforward.

Fix the block size n , and set $R = I(X; Y) + \delta$ for some $\delta > 0$. Let $\{\mathcal{J}_m\}_{m=1}^{\lfloor 2^{nR} \rfloor}$ be an open partition of \mathcal{I} into equi-sized intervals, and let a_m be the midpoint of \mathcal{J}_m . Denote the set of all midpoints by A_n .

Lemma 4. $\lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcup_{\substack{\theta \in A_n \cap \Delta_n^\varepsilon(Y^n) \\ (\theta, \phi) \in u_n(Y^n)}} \{(x^n(\theta, \phi), z^n(\theta, \phi))\} \cap \mathcal{T}_{n, \delta}(P_{XZ}) = \emptyset \right) = 0.$

Proof Outline. For lack of space we only describe the main elements of the proof, skipping some details. Let $V(y^n)$ be the set of indices m such that $a_m \in \Delta_n^\varepsilon(y^n)$, and \mathcal{J}_m intersects with two or more intervals that are projections of a product rectangle in $u_n(y^n)$ onto the θ -axis. By Lemma 2, $|V(y^n)| \leq n|\mathcal{X}|$. Define

$$q_n(y^n) \stackrel{\text{def}}{=} P_{\Theta_1|Y^n} \left(\bigcup_{m \in V(y^n)} \mathcal{J}_m \mid y^n \right)$$

and consider Theorem 1 with a rate $I(X; Y) - \varepsilon_1$. Now, assume to the contrary that $\mathbb{P}(q_n(Y^n) > \varepsilon_2) > \varepsilon_3$ for some fixed $\varepsilon_2 \in (2\varepsilon, 1 - 2\varepsilon)$, $\varepsilon_3 > 0$, i.e., with probability at least ε_3 the distribution of Θ_1 given Y^n has a mass at least ε_2 inside that polynomial sized set of intervals. Then we have (some transitions assuming n large enough)

$$\begin{aligned} n^{-1}I(\Theta_1; Y^n) &= -n^{-1}h(\Theta_1|Y^n) \\ &\geq n^{-1}\varepsilon_3 \left[\varepsilon_2 \log(\varepsilon_2 \cdot 2^{nR} \cdot (n|\mathcal{X}|)^{-1}) \right. \\ &\quad \left. + (1 - 2\varepsilon - \varepsilon_2) \log((1 - 2\varepsilon - \varepsilon_2) \cdot 2^{n(I(X; Y) - \varepsilon_1)}) \right] \\ &\quad + n^{-1}(1 - \varepsilon_3)(1 - 2\varepsilon) \log((1 - 2\varepsilon) \cdot 2^{n(I(X; Y) - \varepsilon_1)}) \\ &= (1 - 2\varepsilon) \cdot I(X; Y) + \delta\varepsilon_2\varepsilon_3 - \varepsilon_1(1 - 2\varepsilon - \varepsilon_2\varepsilon_3) + O(\log n/n) \quad (3) \end{aligned}$$

where we have used the concentration result of Theorem 1 for the inequality transition. Since $\varepsilon, \varepsilon_1$ can be taken arbitrarily small for n large enough, the right-hand-side of (3) can be made larger than $I(X; Y)$, contradicting Lemma 3. Note that this argument is similar in essence to the converse to the channel coding Theorem [12].

We conclude that $q_n(Y^n) \rightarrow 0$ in probability, which loosely speaking means that with high probability, $P_{\Theta_1|Y^n}$ is mostly concentrated on $\Delta_n^\varepsilon(y^n) \setminus \bigcup_{m \in V(y^n)} \mathcal{J}_m$ for large n . Using typicality arguments together with the properties in Lemma 3, this can be shown to imply that with high probability we can find θ in that set together with some ϕ such that $(\theta, \phi) \in u_n(Y^n)$ and $(x^n(\theta, \phi), z^n(\theta, \phi)) \in \mathcal{T}_{n, \delta}(P_{XZ})$. By definition, $\theta \in \mathcal{J}_m$ where \mathcal{J}_m is a subset of some interval which is a projection of a product rectangle in $u_n(Y^n)$. This in turn implies that $x^n(a_m, \phi) = x^n(\theta, \phi)$ and $z^n(a_m, \phi) = z^n(\theta, \phi)$, concluding the proof. \square

We are now ready to describe the compression protocol.

Encoder

- (a) Given the sequence y^n , compute $\Delta_n^\varepsilon(y^n)$ using the recursion (2).
- (b) Out of the $\approx 2^{n\delta}$ intervals $\mathcal{J}_m \subseteq \Delta_n^\varepsilon(y^n)$, find the one with the least index³ m , for which there exists $\phi \in \mathcal{I}$ such that $(x^n(a_m, \phi), z^n(a_m, \phi)) \in \mathcal{T}_{n, \varepsilon}(P_{XZ})$. If no such index exists, arbitrarily set $m = 1$.
- (c) Send the index m to the decoder, which requires a rate of $I(X; Y) + \delta$ bits per source symbol.

³It seems that a random selection should work with high probability, making the process simpler. However, this was not verified.

Decoder

- (a) Initialization: Set $\theta_1 = a_m$, compute $x_1 = \sigma_0(\theta_1)$.
- (b) For any k , predict $\hat{y}_k = x_k$.
- (c) Receive the true y_k via the feedforward link, compute $\theta_{k+1} = T_0(\theta_k, y_k)$ and $x_{k+1} = \sigma_0(\theta_{k+1})$.
- (d) Repeat steps (b)–(c) up to $k = n$.

The compression rate attained by the scheme is $R = I(X; Y) + \delta$. If encoding step (b) is successful then the pair (x^n, z^n) is jointly P_{XZ} -typical, which implies that (x^n, y^n) is jointly P_{XY} -typical. By Lemma 4, when encoding an i.i.d- P_Y sequence Y^n this occurs with probability approaching 1 as $n \rightarrow \infty$. Since the distortion measure is bounded, the expected distortion achieved by the scheme is given by $D = \mathbb{E}_{P_{XY}} d(X, Y) + o(1)$.

The development above holds for any P_Y and $P_{X|Y}$ that satisfy the requirements of Theorem 1. The strict positivity constraint for P_{XY} has a negligible effect, since such distributions can always be approximated arbitrarily via admissible distributions, and the distortion measure is bounded. The second constraint is redundant as it can always be averted by using a variant of the probabilistic dynamical source, as in the channel coding case [9][13]. Hence, we have proved the following result.

Theorem 2. *For any discrete memoryless source and bounded distortion measure, the protocol described above can perform arbitrarily close to the rate distortion function of the source.*

Example 3 (Bernoulli Source and Hamming Distortion). Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $Y \sim \text{Bern}(\frac{1}{2})$, $d(\cdot, \cdot)$ the Hamming distortion measure. The rate distortion function $R_{\text{ff}}(D) = R(D) = 1 - h_b(D)$ is achieved by $X \sim \text{Bern}(\frac{1}{2})$, $Z \sim \text{Bern}(D)$ independent of X , and $Y = X + Z \pmod{2}$. The partitions and mappings are given by

$$\begin{aligned} \Pi_0^0 &= (0, \frac{1}{2}) & \Pi_1^0 &= (\frac{1}{2}, 1) \\ \Pi_0^1 &= (0, 1 - D) & \Pi_1^1 &= (1 - D, 1) \\ T_0(\theta, 0) &= 2\theta(1 - D) \cdot \mathbf{1}_{\Pi_0^0}(\theta) + (2D\theta + 1 - 2D) \cdot \mathbf{1}_{\Pi_1^0}(\theta) \\ T_0(\theta, 1) &= 2\theta D \cdot \mathbf{1}_{\Pi_0^0}(\theta) + (2(1 - D)\theta + 2D - 1) \cdot \mathbf{1}_{\Pi_1^0}(\theta) \\ T_1(\phi, k) &= T_1(\phi) = \frac{\phi}{1 - D} \cdot \mathbf{1}_{\Pi_0^1}(\phi) + \frac{\phi - (1 - D)}{D} \cdot \mathbf{1}_{\Pi_1^1}(\phi) \end{aligned}$$

The mappings and the fundamental sets for $n = 3$ are depicted in Figures 1 and 2.

V Conclusions

A symbolic dynamical system approach to lossy source coding with feedforward was introduced, yielding in particular a conceptually simple and optimal compression protocol for memoryless sources. In this latter case, the construction is dual to the posterior matching feedback communication scheme for memoryless channels. Future work should examine the suggested framework for sources with memory. A reasonable first goal could be the case where the ϕ -component of the dynamical source evolves independently as in the memoryless case, yet generates e.g. a Markovian Z^n .

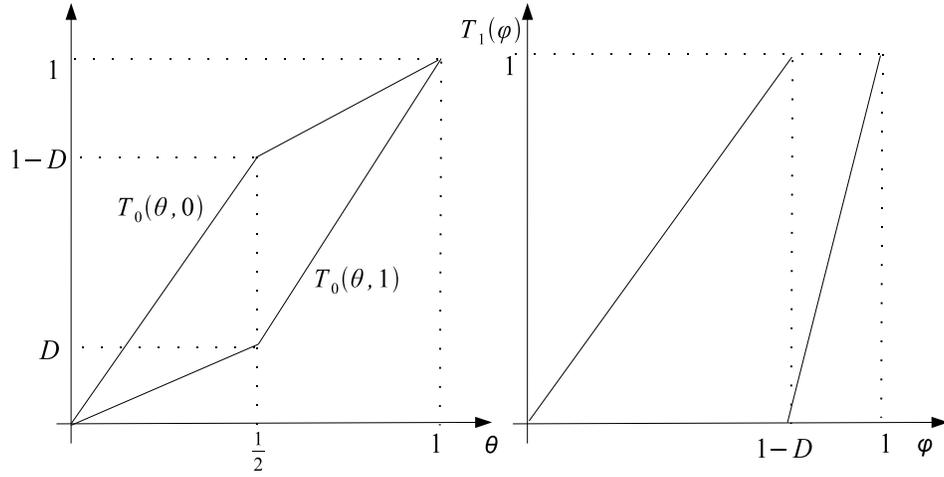


Figure 1: The mapping $T = (T_0, T_1)$ for $\text{Bern}(\frac{1}{2})$ source, Hamming distortion D

1	101				000			
$1-D^3$	100				001			
$1-D^3$	111				010			
D^3-2D^2+1	110				011			
$1-D$	011		000		111		100	
$(1+D)(1-D)^2$	010		001		110		101	
$(1-D)^2$	001	000	011	010	101	100	111	110
$(1-D)^3$	000	001	010	011	100	101	110	111
	$\frac{1}{8(1-D)^2}$	$\frac{1}{4(1-D)}$	$\frac{5-6D}{8(1-D)^2}$	$\frac{1}{2}$	$1-\frac{5-6D}{8(1-D)^2}$	$1-\frac{1}{4(1-D)}$	$1-\frac{1}{8(1-D)^2}$	1

Figure 2: Fundamental sets with $n = 3$ for $\text{Bern}(\frac{1}{2})$ source, Hamming distortion D

References

- [1] B. Vallée, “Dynamical sources in information theory: Fundamental intervals and word prefixes,” *Algorithmica*, vol. 29, pp. 262–306, February 2001.
- [2] M. Halfant, “Analytic properties of Rényi’s invariant density,” *Israel Journal of Mathematics*, vol. 27, no. 1, 1977.
- [3] T. Weissman and N. Merhav, “On competitive prediction and its relation to rate-distortion theory,” *IEEE Trans. Info. Theory*, vol. 49, no. 12, pp. 3185–3194, Dec. 2003.
- [4] R. Venkataramanan and S. Sandeep Pradhan, “Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source,” *Information Theory, IEEE Transactions on*, vol. 53, no. 6, pp. 2154–2179, June 2007.
- [5] E. Martinian and G.W. Wornell, “Source coding with fixed lag side information,” in *Proceedings of the 42th Allerton Conference on Communication, Control, and Computing*.
- [6] S. Sandeep Pradhan, “Source coding with feedforward: Gaussian sources,” in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, June-2 July 2004.
- [7] O. Shayevitz and M. Feder, “Communication with feedback via posterior matching,” in *Proc. of the International Symposium on Information Theory*, 2007.
- [8] O. Shayevitz and M. Feder, “The posterior matching feedback scheme: Capacity achieving and error analysis,” in *Proc. of the International Symposium on Information Theory*, 2008.
- [9] O. Shayevitz and M. Feder, “Optimal feedback communication via posterior matching,” submitted to *IEEE Trans. Info. Theory*, available online at arXiv:0909.4828 [cs.IT].
- [10] M. Iosifescu and C. Kraaikamp, *Metrical Theory of Continued Fractions*, Kluwer Academic Publishers, 2002.
- [11] A. El Gamal and Y.-H Kim, *Lecture Notes on Network Information Theory*, Stanford University and UCSD, 2009.
- [12] T.M. Cover and J.A Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.
- [13] O. Shayevitz, “Posterior matching variants and fixed-point elimination,” in *Proceedings of the 47th Allerton Conference on Communication, Control, and Computing*.