Fundamental Limits of Almost Lossless Analog Compression

Yihong Wu
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: yihongwu@princeton.edu

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*—In Shannon theory, lossless source coding deals with the optimal compression of discrete sources. Compressed sensing is a lossless coding strategy for analog sources by means of multiplication by real-valued matrices. In this paper we study almost lossless analog compression for analog memoryless sources in an information-theoretic framework, in which the compressor is not constrained to linear transformations but it satisfies various regularity conditions such as Lipschitz continuity. The fundamental limit is shown to be the information dimension proposed by Rényi in 1959.**

I. Introduction

The "Bit" is the universal currency in lossless source coding theory [1], where Shannon entropy is the fundamental limit of compression rate for sources modeled by stochastic processes. As probability is concentrated on a set of exponentially small cardinality as blocklength grows, by encoding this subset data compression is achieved if we tolerate a positive, though arbitrarily small, block error probability.

Compressed sensing ([2], [3]) has recently emerged as an approach to lossless encoding of analog sources by real numbers rather than bits. The formulation of the problem is reminiscent of traditional lossless data compression in the following sense:

- Source sparsity (a subset of dimensions is assumed to be zero) is exploited to achieve effective compression by taking fewer number of linear measurements.
- Block error probability, instead of distortion, is the performance benchmark, in contrast to lossy data compression.
- The central problem is to determine how many compressed measurements are sufficient / necessary for recovery with vanishing block error probability as blocklength tends to infinity ([2], [3], [4]).
- Random coding is employed to show the existence of "good" linear encoders.

On the other hand, there are also significantly different ingredients in compressed sensing that differ from information theoretic setups, such as:

- Sources are not modeled probabilistically, and the fundamental limits are on a worst-case basis rather than on average. Moreover, block error probability is with respect to the distribution of the encoding random matrices.
- Real-valued sparse vectors are encoded by real numbers instead of bits.
- The encoder is confined to be linear, while generally in information-theoretical problems such as lossless source coding we have the freedom to choose the best possible coding scheme.

Departing from the compressed sensing literature, we study the fundamental limits of lossless source coding for real-valued memoryless sources within an information theoretic setup.

- Sources are modeled by random processes. This method is more flexible to describe source redundancy, which encompasses, but is not limited to, sparsity. For example, a mixed discrete-continuous distribution is suitable for characterizing linearly sparse vectors, i.e. those with a number of nonzero components proportional to the blocklength with high probability and whose nonzero components are distributed according to some continuous distribution on the support.
- Block error probability is evaluated with respect to the source distribution.
- Encoders/decoders are subject to more general regularity conditions than linearity.

It turns out that under several important regularity conditions, the fundamental limit is the *information dimension of the source*, an information measure for random vectors in Euclidean space proposed by Alfréd Rényi in 1959. It characterizes the rate of growth of the information given by successively finer discretizations of the space. Although a fundamental information measure, it is far less well-known than either the Shannon entropy or the Rényi entropy. Rényi showed in [5] that under certain conditions for an absolutely continuous n-dimensional random vector the information dimension is n. Hence he remarked that "*... the geometrical (or topological) and information-theoretical concepts of dimension coincide for absolutely continuous probability distributions*". However, the operational role of Rényi information dimension has not been addressed before except in the work of Kawabata and Dembo [6], which relates it to the rate-distortion function. In this paper we give a new operational characterization of Rényi information dimension as the fundamental limit of almost lossless data compression for analog sources under various regularity constraints of the encoder/decoder.

The rest of the paper is organized as follows. Section II gives an overview of Rényi information dimension and a new interpretation in terms of entropy rate. Section III states the main definitions and results in the paper, and discusses its implications. Section IV concludes the paper. Due to space limitations, all proofs are referred to [7].

II. Rényi Information Dimension

In [5], Rényi defined the *information dimension of a probability distribution* as follows:

Definition 1 (Information Dimension): Let X be a real-valued random variable. Denote for a positive integer m the quantized version of X:

$$\langle X\rangle_m = \frac{\lfloor mX\rfloor}{m}. \tag{1}$$

Define

$$\underline{d}(X) = \liminf_{m\to\infty} \frac{H\left(\langle X\rangle_m\right)}{\log m} \tag{2}$$

and

$$\overline{d}(X) = \limsup_{m\to\infty} \frac{H\left(\langle X\rangle_m\right)}{\log m}, \tag{3}$$

where $\underline{d}(X)$ and $\overline{d}(X)$ are called *lower* and *upper information dimensions* of X respectively. If $\underline{d}(X) = \overline{d}(X)$, the common value is called the *information dimension* of X, denoted by $d(X)$.

Definition 1 can be readily extended to random vectors, where the floor function $\lfloor\cdot\rfloor$ is taken componentwise. With Shannon entropy replaced by Rényi entropy of order α in (2 – 3), the generalized notion of dimension of order α is defined similarly, denoted by $\overline{d}_\alpha(X)$, $\underline{d}_\alpha(X)$ and $d_\alpha(X)$.

Note that the lower and upper information dimension of a random variable are finite if and only if the mild condition $\mathbb{E}\log(|X|+1) < \infty$ is satisfied, as the following proposition shows:

Proposition 1: If $\mathbb{E}\log(|X|+1) < \infty$, then

$$0 \le \underline{d}(X) \le \overline{d}(X) \le 1. \tag{4}$$

If $\mathbb{E}\log(|X|+1) = \infty$, then

$$\underline{d}(X) = \overline{d}(X) = \infty. \tag{5}$$

It can be shown that to calculate the information dimension, it is sufficient to restrict to the exponential subsequence $m = 2^l$ in (2) and (3). This observation leads to an entropy-rate interpretation of the information dimension. Define the quantization operator $[\cdot]_l = \langle\cdot\rangle_{2^l}$ and $(x)_l = 2^l([x]_l - [x]_{l-1})$. Consider $X \in [0,1)$ a.s. The binary expansion of X can be written as

$$X(\omega) = \sum_{j=1}^{\infty} (X)_j(\omega) 2^{-j}, \tag{6}$$

where each $(X)_j$ is a binary random variable. Note that $[X]_i(\omega) = \sum_{j=1}^{i}(X)_j(\omega)2^{-j}$ is in one-to-one correspondence with $((X)_1, \ldots, (X)_i)$, therefore

$$\underline{d}(X) = \liminf_{i\to\infty} \frac{H\left((X)_1, \ldots, (X)_i\right)}{i}, \tag{7}$$

$$\overline{d}(X) = \limsup_{i\to\infty} \frac{H\left((X)_1, \ldots, (X)_i\right)}{i}. \tag{8}$$

In general, if the information dimension of a real-valued random variable $d(X)$ is finite, it coincides with the *entropy rate* of of any M-ary expansion of the fractional part of X.

By the Lebesgue Decomposition Theorem [8], any probability distribution can be uniquely represented as the mixture of a discrete, an absolutely continuous and a singular (with respect to Lebesgue measure) probability measure. The information dimension for discrete-continuous mixture can be determined as follows:

Theorem 1 ([5]): Let X be a random variable such that $H(\lfloor X\rfloor)$ is finite. Assume the distribution of X can be represented as $\nu = (1-\rho)\nu_{\mathrm{d}} + \rho\nu_{\mathrm{c}}$, where ν_{d} is a discrete measure, ν_{c} is an absolutely continuous measure and $0 \le \rho \le 1$. Then

$$d(X) = \rho. \tag{9}$$

Therefore, when X has a discrete-continuous mixed distribution, the information dimension of X is exactly the weight of the continuous part. When the distribution of X has a singular component, its information dimension does *not* admit a simple formula in general. In fact [5] if X has a singular distribution, it is possible that $\underline{d}(X) < \overline{d}(X)$. However, for the important class of *self-similar* singular distributions, the information dimension can be explicitly determined [9], [6]. For example, the Cantor distribution [10] has information dimension $\log_3 2$.

III. Definitions and Main results

This section presents a unified framework for lossless data compression and our main results in the form of coding theorems under various regularity conditions.

A. Lossless Data Compression

Let the source $\{X_i : i \in \mathbb{N}\}$ be a stochastic process on $(\mathcal{X}^{\mathbb{N}}, \mathcal{F}^{\otimes\mathbb{N}})$, with $\mathcal{X}$ denoting the source alphabet and $\mathcal{F}$ a σ-algebra over $\mathcal{X}$. Let $(\mathcal{Y}, \mathcal{G})$ be a measurable space, where $\mathcal{Y}$ is called the *code alphabet*. The main objective of lossless data compression is to find efficient representations for source realization $x^n \in \mathcal{X}^n$ by a string $y^k \in \mathcal{Y}^k$.

Definition 2: A (n,k)-code for $\{X_i : i \in \mathbb{N}\}$ over the code space $(\mathcal{Y}, \mathcal{G})$ is a pair of mappings:

1) Encoder: $f_n : \mathcal{X}^n \to \mathcal{Y}^k$ that is measurable relative to $\mathcal{F}^n$ and $\mathcal{G}^k$.
2) Decoder: $g_n : \mathcal{Y}^k \to \mathcal{X}^n$ that is measurable relative to $\mathcal{G}^k$ and $\mathcal{F}^n$.

For an (n, k)-code, its *codebook* is denoted as the collection of codewords corresponding to source realizations that are correctly decoded, i.e.,

$$C_n = \{f_n(x^n) : x^n \in \mathcal{X}^n, g_n(f_n(x^n)) = x^n\} \subset \mathcal{Y}^k. \quad (10)$$

And the block error probability is $\mathbb{P}\{g_n(f_n(X^n)) \neq X^n\}$.

The fundamental limit in lossless source coding is:

Definition 3 (Lossless Data Compression): Let $\{X_i : i \in \mathbb{N}\}$ be a stochastic process on $(\mathcal{X}^{\mathbb{N}}, \mathcal{F}^{\otimes\mathbb{N}})$. Define the minimum ϵ-achievable rate $r(\epsilon)$ to be the infimum of $r > 0$ such that there exists a sequence of $(n, \lfloor rn \rfloor)$-codes (f_n, g_n) over the code space $(\mathcal{Y}, \mathcal{G})$, such that

$$\mathbb{P}\{g_n(f_n(X^n)) \neq X^n\} \leq \epsilon \quad (11)$$

holds for all sufficiently large n.

If $\mathcal{X}$ is countable and $\mathcal{Y}$ is finite, equipped with discrete σ-algebras $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$ respectively, the measurability of encoder and decoder are trivially satisfied. This is one of the reasons why in conventional lossless coding theory we only consider how to choose a subset of source symbols of small cardinality and large probability, regardless of how they are mapped to each codeword as long as the mapping is injective. The fundamental limit of lossless compression is

Theorem 2 ([1]): Let $\mathcal{X}$ be countable and $\mathcal{Y}$ finite. $\{X_i : i \in \mathbb{N}\}$ is an i.i.d. $\mathcal{X}$-valued process with common distribution P. Then for any finite code alphabet

$$r(\epsilon) = \begin{cases} \frac{\log|\mathcal{X}|}{\log|\mathcal{Y}|}, & \epsilon = 0 \\ \frac{H(P)}{\log|\mathcal{Y}|}, & 0 < \epsilon < 1, \\ 0, & \epsilon = 1. \end{cases} \quad (12)$$

Using codes over an infinite alphabet, any discrete source can be compressed with zero rate and zero block error probability.

Proposition 2: Let $\mathcal{X}$ be countable and $\mathcal{Y}$ countably infinite. $\{X_i : i \in \mathbb{N}\}$ is an $\mathcal{X}$-valued random process. Then for $0 \leq \epsilon \leq 1$,

$$r(\epsilon) = 0. \quad (13)$$

B. Lossless Analog Compression with Regularity Conditions

In this subsection we consider the problem of encoding analog sources with analog symbols, that is, $(\mathcal{X}, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ and $(\mathcal{Y}, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$ or $([0,1], \mathcal{B})$ if *bounded* encoders are required, where $\mathcal{B}$ denotes the Borel σ-algebra. As in Proposition 2, we see that, without constraints on the encoding or decoding method, zero rate is achievable even for zero block error probability, because the cardinality of $\mathbb{R}^n$ is the same for any n [11]. This conclusion holds even if we require the encoder/decoder to be Borel measurable, because according to Kuratowski's theorem [12, Remark (i), p. 451] every uncountable standard Borel space is isomorphic to $([0,1], \mathcal{B})$. Therefore a single real number has the capability of encoding a real vector, or even a real sequence, with a coding scheme that is both *universal* and *deterministic*.

However, the rich structure of $\mathbb{R}$ equipped with a metric topology (*e.g.*, that induced by Euclidean distance) enables us to probe the problem further. If we seek the fundamental limits of not only *lossless* coding but *"graceful" lossless* coding, the result is not trivial anymore. This is our primary goal throughout the paper. In this spirit, our various definitions share the basic information-theoretic setup where a random vector is encoded with a function $f_n : \mathbb{R}^n \to \mathbb{R}^{\lfloor Rn \rfloor}$ and decoded with $g_n : \mathbb{R}^{\lfloor Rn \rfloor} \to \mathbb{R}^n$ with $R \leq 1$ such that f_n and g_n satisfy certain regularity conditions and the probability of incorrect reproduction vanishes as $n \to \infty$.

Regularity in encoder and decoder is imposed for the sake of both less complexity and more robustness. For example, although the surjection g from $[0,1]$ to $\mathbb{R}^n$ is capable of lossless encoding, its irregularity requires specifying *uncountably* many real numbers to determine this mapping. However, if g is continuous, the Stone-Weierstrass theorem states that g can be uniformly approximated by polynomials, hence *countably* many coefficients will be sufficient to characterize g. Moreover, regularity in encoder/decoder is crucial to guarantee noise resilience of the coding scheme. This may be beneficial even in the discrete world [13]. For instance, if the encoder (decoder resp.) is L-Lipschitz with respect to the ℓ_0 distance, then changing one entry of the data (codeword resp.) will result in changes of at most L positions in the encoder (decoder resp.) output, which is a desirable feature in some source coding applications.

Definition 4: Let $\{X_i : i \in \mathbb{N}\}$ be a stochastic process on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$. Define the minimum ϵ-achievable rate to be the infimum of $R > 0$ such that there exists a sequence of $(n, \lfloor Rn \rfloor)$-codes (f_n, g_n) , such that [1]

$$\limsup_{n\to\infty} \mathbb{P}\{g_n(f_n(X^n)) \neq X^n\} \leq \epsilon \quad (14)$$

and the encoder f_n and decoder g_n are constrained according to Table I. Except for linear encoding where $\mathcal{Y} = \mathbb{R}$, it is assumed that $\mathcal{Y} = [0,1]$.

TABLE I
REGULARITY CONDITIONS OF ENCODER/DECODERS AND CORRESPONDING NOTATIONS FOR MINIMUM ϵ-ACHIEVABLE RATES.

Encoder	Decoder	Minimum ϵ-achievable rate
Borel	Continuous	$\mathsf{R}_0(\epsilon)$
Continuous	Continuous	$\tilde{\mathsf{R}}(\epsilon)$
Linear	Continuous	$\mathsf{R}^*(\epsilon)$
Borel	Lipschitz	$\mathsf{R}(\epsilon)$
Borel	Δ-stable	$\overline{\mathsf{R}}(\epsilon, \Delta)$

In Definition 4 we have used the following definitions:

Definition 5 ((L, Δ)-stable): Let (U, d_U) and (V, d_V) be metric spaces and $T \subset U$. $g : U \to V$ is called (L, Δ)-stable on T if for all $x, y \in T$

$$d_U(x, y) \leq \Delta \Rightarrow d_V(g(x), g(y)) \leq L\Delta. \quad (15)$$

And we say g is Δ-stable if g is $(1, \Delta)$-stable.

[1] Definition 14 applies to $0 < \epsilon \leq 1$; for $\epsilon = 0$, we require that $g_n(f_n(X^n)) = X^n$ with probability 1 for all but a finite number of n.

An even stronger condition than (L,Δ)-stability is Lipschitz continuity, in that we have the following characterization of Lipschitz functions as in Proposition 3.

Definition 6 (Hölder and Lipschitz continuity): Let (U,d_U) and (V,d_V) be metric spaces. $g : U \to V$ is called (L,γ)-*Hölder continuous* if there exists $L,\gamma \geq 0$ such that for any $x,y \in U$,

$$d_V(g(x),g(y)) \leq L d_U(x,y)^\gamma. \tag{16}$$

g is called L-*Lipschitz* if g is $(L,1)$-Hölder continuous.

Proposition 3: $g : U \to V$ is L-Lipschitz if and only if g is (L,Δ)-stable for every $\Delta > 0$.

We first give a general theorem about the ordering of various minimum ϵ-achievable rates introduced in Definition 4:

Theorem 3:

$$0 = \mathsf{R}_0(\epsilon) \leq \tilde{\mathsf{R}}(\epsilon) \leq \mathsf{R}^*(\epsilon) \leq \mathsf{R}(\epsilon) \tag{17}$$

holds for any sources and any $0 < \epsilon \leq 1$.

The first equality, corresponding to the case where the decoder is constrained to be continuous but the encoder is only assumed to be Borel measurable, is a consequence of the Hahn-Mazurkiewicz theorem in topology [14]. The last inequality is a surprising result, which asserts that a Lipschitz constraint at the decompressor results in less efficient compression than a linearity constraint at the compressor. The rest of the inequalities in (17) follow from Definition 4.

Now we proceed to give results for each of the minimum ϵ-achievable rates for memoryless sources. In view of compressed sensing theory, it is interesting to consider the case where the encoder is restricted to be linear.

Theorem 4 (Linear encoding: general achievability): For memoryless sources,

$$\mathsf{R}^*(\epsilon) \leq \hat{d}(X) = \lim_{\alpha \uparrow 1} \underline{d}_\alpha(X) \tag{18}$$

holds for all $0 < \epsilon < 1$. Moreover,

1) For Lebesgue-a.e. linear encoder, block error probability ϵ is achievable with a uniformly continuous decoder.
2) The decoder can be chosen to be β-Hölder continuous for all $0 < \beta < \frac{R-\hat{d}(X)}{R}$, where $R > \hat{d}(X)$ is the compression rate.

Theorem 5 (Linear encoding: discrete-continuous mixture): For memoryless sources, if X has a discrete-continuous mixed distribution, then for $0 < \epsilon < 1$,

$$\mathsf{R}^*(\epsilon) = d(X). \tag{19}$$

Next we drop the restriction that the encoder is linear, allowing very general encoding rules. Requiring both the encoder and the decoder to be continuous, we give two achievability results as follows:

Theorem 6 (Continuous encoder and decoder): For memoryless sources and $0 < \epsilon < 1$,

1)
$$\tilde{\mathsf{R}}(\epsilon) \leq \lim_{\alpha \uparrow 1} \underline{d}_\alpha(X). \tag{20}$$

2) Furthermore, if the source has a discrete-continuous mixed distribution, then

$$\tilde{\mathsf{R}}(\epsilon) \leq d(X). \tag{21}$$

Theorems 7 - 9 deal with Lipschitz decoding in Euclidean spaces.

Theorem 7 (Lipschitz decoding: general converse): For memoryless sources, and $0 < \epsilon < 1$, if $\overline{d}(X) < \infty$, then

$$\mathsf{R}(\epsilon) \geq \overline{d}(X). \tag{22}$$

Theorem 8 (Lipschitz decoding: discrete/continuous mixture): For memoryless sources, if X has a discrete-continuous mixed distribution, then for $0 < \epsilon < 1$,

$$\mathsf{R}(\epsilon) = d(X). \tag{23}$$

For sources with a singular distribution, in general there is no simple answer due to their fractal nature. For an important class of singular measures, namely self-similar measures generated from i.i.d. digits (e.g. generalized Cantor distributions [10]), the information dimension turns out to be the fundamental limit for lossless compression with Lipschitz decoder, which can be constructed using finitary homomorphism theorems in ergodic theory.

Theorem 9 (Lipschitz decoding: self-similar measures): Let the distribution of X be a self-similar measure generated by i.i.d. M-ary digits with common distribution P. Then for $0 < \epsilon < 1$,

$$\mathsf{R}(\epsilon) = d(X) = \frac{H(P)}{\log M}. \tag{24}$$

Moreover, if P is equiprobable on its support, then

$$\mathsf{R}(0) = d(X). \tag{25}$$

Remark 1: The proofs of Theorems 3 – 7 all hinge upon the lossless Minkowski dimension compression theory developed in [7], which is a counterpart of the conventional lossless data compression.

Remark 2: As an example, we consider the setup in Theorem 9 with $M = 3$ and $P = \{p, 0, q\}$, where $p + q = 1$. The associated invariant set is the middle third Cantor set C [10] and X is supported on C. The distribution of X, denoted by μ, is called the generalized Cantor distribution [15]. In the ternary expansion of X, each digit is independent with probability p being 0 and q being 2. Then by Theorem 9, for any $\epsilon \in (0,1)$, $\mathsf{R}(\epsilon) = \frac{h_2(p)}{\log 3}$. Furthermore, when $p = 1/2$, μ coincides with the 'uniform' distribution on C, i.e., the standard Cantor distribution. Hence (25) implies a stronger result that $\mathsf{R}(0) = \log_3 2 \approx 0.63$, i.e., information dimension achieves exact lossless compression with Lipschitz continuous decompressor.

For stable decoding the fundamental limit is given by the following tight result:

Theorem 10 (Δ-Stable decoding): Let the underlying metric be the ℓ_∞ distance. Then for memoryless sources, and $0 < \epsilon < 1$,

$$\limsup_{\Delta \downarrow 0} \overline{\mathsf{R}}(\epsilon, \Delta) = \overline{d}(X), \tag{26}$$

that is, the minimum ϵ-achievable rate such that for all sufficiently small Δ there exists a Δ-stable coding strategy is given by $\overline{d}(X)$.

C. Connections with Compressed Sensing

Robust reconstruction is of great importance in compressed sensing, since *noise resilience* is an indispensable property for decompressing sparse signals from real-valued measurements. For example, consider the following robustness result:

Theorem 11 ([16]): Suppose we wish to recover a vector $x_0 \in \mathbb{R}^n$ from k noisy compressed linear measurements $y = \mathbf{A}x_0 + e$, where $\mathbf{A} \in \mathbb{R}^{k \times n}$, $e, y \in \mathbb{R}^k$ and $\|e\|_2 \le \epsilon$. Let $x^\#$ be a solution of the following ℓ_1-regularization problem

$$\begin{aligned} \min \quad & \|x\|_1 \\ \text{s.t.} \quad & \|\mathbf{A}x - y\|_2 \le \epsilon. \end{aligned} \tag{27}$$

Assume that $\|x_0\|_0 = S$ satisfies $\delta_{3S} + 3\delta_{4S} \le 2$, where δ_S is the S-restricted isometry constant δ_S of matrix $\mathbf{A}$, defined as the smallest positive number such that

$$(1 - \delta_S)\|u\|_2^2 \le \|\mathbf{A}u\|_2^2 \le (1 + \delta_S)\|u\|_2^2 \tag{28}$$

holds for all $T \subset \{1, \ldots, n\}$ with $|T| \le S$ and for all u in $\mathbb{R}^n$ supported on T. Then

$$\left\|x^\# - x_0\right\|_2 \le C_S \epsilon, \tag{29}$$

where C_S may only depend on δ_{4S}.

By Theorem 11, using (27) as the decoder, the ℓ_2 norm of the decoding error is upper bounded *proportionally* to the ℓ_2 norm of the noise.

In our framework a stable or Lipschitz continuous coding scheme also implies *robustness* with respect to noise added at the input of the decompressor, which could result from quantization, finite wordlength or other inaccuracies. For example, suppose that the encoder output $y^k = f_n(x^n)$ is quantized by a q-bit uniform quantizer, resulting in $\tilde{y}^k$. With a 2^{-q}-stable coding strategy, we can use the following decoder: denote the following nonempty set

$$\mathcal{D}(\tilde{y}^k) = \left\{ z^k \in C_n : \left\|z^k - \tilde{y}^k\right\|_\infty \le 2^{-q} \right\}. \tag{30}$$

where C_n denotes the codebook. Pick any z^k in $\mathcal{D}(\tilde{y}^k)$ and output $\hat{x}^n = g_n(z^k)$. Then by the stability of g_n, we have

$$\|\hat{x}^n - x^n\|_\infty \le 2^{-q}, \tag{31}$$

i.e., each component in the decoder output will suffer at most the inaccuracy of the decoder input. Similarly, an L-Lipschitz coding scheme with respect to ℓ_∞ distance incurs an error no greater than $L2^{-q}$.

IV. Concluding Remarks

In this paper we have proposed an information theoretic framework for lossless analog compression of analog sources under regularity conditions of the coding schemes. It could also be viewed as a probabilistic dimension reduction problem with smooth embedding. In this framework, obtaining fundamental limits requires tools quite different from those used in traditional information theoretical problems, calling for machineries from dimension theory, geometric measure theory and ergodic theory. As seen in Theorems 4 – 10, Rényi's information dimension plays a fundamental role in the associated coding theorems.

In the important case of discrete-continuous mixed sources, we have shown that the fundamental limit is Rényi information dimension, which coincides with the weight on the continuous part in the source distribution. In the memoryless case, this corresponds to the percentage of analog symbols in the source realization. This might suggest that the mixed discrete-continuous nature of the source is of fundamental importance in the analog compression framework; sparsity is just one manifestation of a mixed distribution.

Acknowledgment

The authors would like to thank Mung Chiang and Jouni Luukkainen for stimulating discussions.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379 – 423, 623–656, 1948.

[2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 – 1306, April 2006.

[3] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, February 2006.

[4] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *Proceedings of 2007 IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[5] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1 – 2, March 1959.

[6] T. Kawabata and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1564 – 1572, 1994.

[7] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," submitted to IEEE Transactions on Information Theory, 2009.

[8] E. Çinlar, "Lecture Notes in Probability Theory: ORF551 Princeton University," 2007.

[9] K. Falconer, *Techniques in Fractal Geometry*. John Wiley, 1997.

[10] ——, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. John Wiley & Sons, Ltd, 2003.

[11] P. Halmos, *Naive Set Theory*. Princeton, NJ: D. Van Nostrand Company, 1960.

[12] K. Kuratowski, *Topology*. New York: Academic Press, 1966, vol. 1.

[13] A. Montanari and E. Mossel, "Smooth compression, Gallager bound and nonlinear sparse graph codes," in *Proceedings of 2008 IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[14] J. G. Hocking and G. S. Young, *Topology*. Addison-Wesley Publishing company, 1961.

[15] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge University Press, 1999.

[16] E. Candés, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59(8), pp. 1207–1223, August 2006.