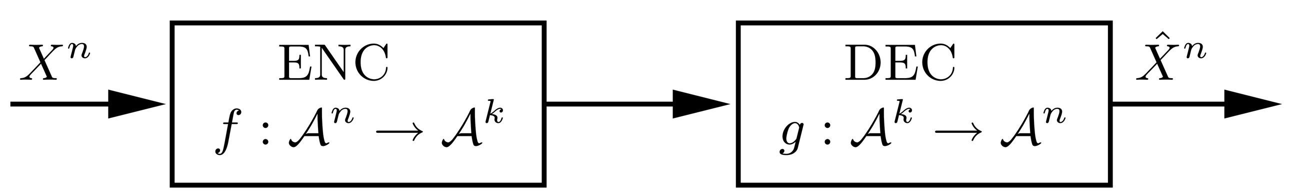
# Fundamental Limits of Almost Lossless Analog Compression


Yihong Wu and Sergio Verdú

Department of Electrical Engineering, Princeton University, Princeton NJ, USA


## Introduction

**Date Compression: from discrete to continuous alphabet**




- Memoryless source with alphabet $\mathcal{A}$.
- Block error probability $\epsilon_n = \mathbb{P}\{g(f(X^n)) \neq X^n\}$;
- Dimension compression rate: $R = \frac{k}{n}$.
- Fundamental limit:
  - Finite alphabet: $\frac{H(X)}{\log|\mathcal{A}|}$, by Shannon's theorem;
  - Continuum alphabet: 0 (achievable even for zero error probability), by a cardinality argument.
  - Problem: irregularity of compressor/decompressor.

**Motivation from Compressed Sensing (CS)**

**CS as an analog compression paradigm:**

- Exploits source redundancy: sparsity.
- Achieves vanishing block error probability.
- Key question: how many measurements are sufficient.
- Compressor: linearity for low complexity.
- Decompressor: Lipschitz continuity for robust reconstruction with respect to noise.

**Comparison:**

| | Data Compression | Compressed Sensing | Analog Compression |
|---|---|---|---|
| Source | Random | Deterministic | Random |
| Analysis | Average | Worst-case | Average |
| Codes | Bits | Reals | Reals |
| Compressor | — | Linear | Regularity conditions |
| Decompressor | — | Lipschitz | Regularity conditions |
| Goal | Vanishing block error probability | | |
| Randomness | Source | Matrix | Source |
| Fundamental limit | $H$ | $2 \times$ sparsity | ? |

**Analogy:**

compressed sensing $\leftrightarrow$ coding theory

analog compression $\leftrightarrow$ information theory

## Almost Lossless Analog Compression

**Definition:**

- $\{X_i : i \in \mathbb{N}\}$: a stochastic process on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$.
- Minimum $\epsilon$-achievable rate: the infimum of $R > 0$ such that there exists a sequence of $(n, \lfloor Rn \rfloor)$-codes $(f_n, g_n)$ such that
  -
  $$\limsup_{n\to\infty} \mathbb{P}\{g_n(f_n(X^n)) \neq X^n\} \leq \epsilon$$
  - $f_n$ and $g_n$ are constrained according to the following table:

**Regularity conditions**

| Compressor | Decompressor | Minimum $\epsilon$-achievable rate |
|---|---|---|
| Borel | Continuous | $\mathsf{R}_0(\epsilon)$ |
| Continuous | Continuous | $\mathsf{R}(\epsilon)$ |
| Linear | Continuous | $\mathsf{R}^*(\epsilon)$ |
| Borel | Lipschitz | $\mathsf{R}(\epsilon)$ |
| Borel | $\Delta$-stable | $\overline{\mathsf{R}}(\epsilon, \Delta)$ |

**Theorem 1 (General ordering of achievable rates)**

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

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

**Surprising result:** Lipschitz decompression is always harder to achieve than linear compression.

## Rényi 's Information Dimension

**Definition:**

$$d(X) = \lim_{m\to\infty} \frac{H(\langle X \rangle_m)}{\log m}.$$

where

- $X$ is a real-valued random variable.
- $\langle X \rangle_m \triangleq \frac{\lfloor mX \rfloor}{m}$ is the quantized version of $X$ of accuracy $\frac{1}{m}$.
- $\underline{d}(X)$ and $\overline{d}(X)$: lower and upper information dimension, corresponding to lim inf and lim sup resp.
- $d_\alpha(X)$: information dimension of order $\alpha$, replacing Shannon entropy $H$ by Rényi entropy $H_\alpha$.

**Properties:**

- If $\mathbb{E}\log(|X|+1) < \infty$, then
  $$0 \leq \underline{d}(X) \leq \overline{d}(X) \leq 1.$$
  If $\mathbb{E}\log(|X|+1) = \infty$, then
  $$\underline{d}(X) = \overline{d}(X) = \infty.$$
- Invariant under affine transformation (but not under arbitrary bijection).
- If $d(X)$ is finite, it coincides with the entropy rate of the binary expansion of the fractional part of $X$.

**Rényi's Theorem [R59]** If $d(X)$ is finite, and

$$P_X = (1-\rho)P_X^{\mathrm{d}} + \rho P_X^{\mathrm{c}},$$

where $P_X^{\mathrm{d}}$ is a discrete distribution and $P_X^{\mathrm{c}}$ is an absolutely continuous distribution, then

$$d(X) = \rho.$$

**Note:** Discrete-Continuous mixtures are good statistical models for linearly sparse sources.

## Coding Theorems for Memoryless Analog Sources

**Linear compression**

**Theorem 2** *For memoryless sources, if $X$ has a discrete-continuous mixed distribution, then*

$$\mathsf{R}^*(\epsilon) = d(X), \quad \forall\, 0 < \epsilon < 1.$$

*Moreover, the decompressor can be chosen to be Lipschitz.*

**Theorem 3 (General Achievabilty)** *For memoryless sources,*

$$\mathsf{R}^*(\epsilon) \leq \hat{d}(X) = \lim_{\alpha\uparrow 1} \underline{d}_\alpha(X), \quad \forall\, 0 < \epsilon < 1.$$

*Moreover,*

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

**Proof techniques:**

- Converse: Steinhaus' theorem.
- Achievability: construction of linear compressors via projection on random subspaces.

**Lipschitz decompression**

**Theorem 4** *For memoryless sources, if $X$ has a discrete-continuous mixed distribution, then*

$$\mathsf{R}(\epsilon) = d(X), \quad \forall\, 0 < \epsilon < 1.$$

**Theorem 5** *Let the distribution of $X$ be a self-similar measure generated by i.i.d. $M$-ary digits with common distribution $P$. Then*

-
$$\mathsf{R}(\epsilon) = d(X) = \frac{H(P)}{\log M}, \quad \forall\, 0 < \epsilon < 1.$$
- *Moreover, if $P$ is equiprobable on its support, then the above holds even for $\epsilon = 0$.*

**Theorem 6 (General converse)** *For memoryless sources, if $\overline{d}(X) < \infty$, then*

$$\mathsf{R}(\epsilon) \geq \overline{d}(X), \quad \forall\, 0 < \epsilon < 1.$$

**Proof techniques:**

- Converse: Minkowski dimension is nondecreasing under Lipschitz mappings.
- Achievability: finitary coding results of Bernoulli shifts from ergodic theory.

**Stable decompression**

**Theorem 7** *Let the underlying metric be the $\ell_\infty$ distance. Then for memoryless sources,*

$$\limsup_{\Delta\downarrow 0} \overline{\mathsf{R}}(\epsilon, \Delta) = \overline{d}(X), \quad \forall\, 0 < \epsilon < 1.$$

## Conclusion

- An information theory for compressed sensing;
- New operational characterizations for information dimension.

## References


**R59** Alfréd Rényi, "On the dimension and entropy of probability distributions", *Acta Mathematica Hungarica*, vol. 10, no. 1 – 2, March 1959.

**WV09a** Yihong Wu and Sergio Verdú, "Fundamental Limits of Almost Lossless Analog Compression," *IEEE International Symposium on Information Theory*, Seoul, Korea, June 2009.

**WV09b** Yihong Wu and Sergio Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression", *submitted to IEEE Transactions on Information Theory*, 2009.