The Delay-Redundancy Tradeoff in Lossless Source Coding

Ofer Shayevitz

UCSD

ITA Workshop, February 2009

Joint work with: E. Meron, M. Feder and R. Zamir (TAU)

Background

- Discrete Memoryless Source (DMS) P
- Lossless coding scheme
- Average end-to-end delay constraint *d*
- The (average) r*edundancy* $\mathcal{R}$ - gap between the average code-length and the entropy $H(P)$
- Traditionally – *dictionary* based encoders, *delay* identified with *block/phrase length*
  - B2V codes (e.g. Huffman): $\mathcal{R} = O(d^{-1})$ [classic][Szpankowski '00]
  - V2B codes (e.g. Tunstall): $\mathcal{R} = O(d^{-1})$ [Savari `97]
  - V2V codes : $\mathcal{R} = O(d^{-\frac{4}{3}})$ [Khodak `69][Bugeaud et al]
- The decay is polynomial!
  - Holds for the more stringent maximal delay constraint

Background – cont.

- Idealized arithmetic coding
  - Attains zero asymptotic redundancy
  - The maximal delay is unbounded
  - However, the average delay is bounded!
    [Gallager`91],[Shayevitz et al `06]
- The apparent disparity is due to
  - Dictionary encoders are resource-oriented and not redundancy-oriented
  - Definition of the delay

Question: What is the redundancy incurred by imposing a <u>maximal</u> end-to-end delay constraint?

Setting

- A DMS P over an alphabet $\mathcal{X}$
- Sequentially emitting symbols $X_1, X_2, \ldots$
- *Encoder* $\mathcal{E}$ *:*
  - A sequence of mappings $\mathcal{E} = \{\mathcal{X}^n \mapsto \{0,1\}^*\}_{n=1}^{\infty}$
  - *Causal* – $\mathcal{E}(x^n)$ is a prefix of $\mathcal{E}(x^n y)$
  - *Integrity property* – $\mathcal{E}(x^n)$ is the maximal common prefix of $\{\mathcal{E}(x^n y)\}_{y \in \mathcal{X}}$
  - Meets a (maximal) *delay constraint d* if $\mathcal{E}(x^n)$ uniquely determines x^{n-d} for any $x^n \in \mathcal{X}^n$, $n > d$

- Lossless encoder, decoder is implicitly defined

Setting – cont.

- *The* (average, asymptotic) *redundancy*

$$\mathcal{R}_{\mathcal{E}}(P) = \limsup_{n\to\infty} \frac{1}{n}\,\mathbb{E}\left(|\mathcal{E}(X^n)|\right) - H(P)$$

- The *redundancy-delay* function for the source P

$$\mathcal{R}(d,P) = \inf_{\mathcal{E}\in\mathcal{F}_d} \mathcal{R}_{\mathcal{E}}(P)$$

where $\mathcal{F}_d$ is the family of all encoders satisfying a delay constraint *d*

Main Results

- Explicit upper bounds on the redundancy-delay function, via a modified arithmetic coder
- Redundancy decays *exponentially* with delay!
  - The encoder does not "reset" after emitting bits
  - The state is always past dependent
- Provides a lower bound on the redundancy-delay exponent

$$E_{rd}(P) = \lim_{d \to \infty} -\frac{1}{d} \log \mathcal{R}(d, P)$$

- Tighter lower bound via typicality
- Upper bound on $E_{rd}(P)$ for almost all sources

Preliminaries

Interval-Mapping Encoders

- "Sufficient" in terms of delay-redundancy tradeoff
- Growing source sequences x^n are mapped into disjoint shrinking intervals $I(x^n) \subseteq [0,1)$

$$I(x^n y) \subset I(x^n)\,, \qquad I(x^n y) \cap I(x^n z) = \phi \quad y \neq z$$

- Time-varying arithmetic coding
- $\mathcal{E}(x^n)$ is the binary sequence repres$I(x^n)$ng the minimal binary interval containing

The Forbidden Points Constraint

- For any encoder interval $I(x^n)$ there exists a countable set $S_{I(x^n)} \subset I(x^n)$ of *forbidden points*
- The encoder meets a delay constraint *d* if and only if

$$I(x^{n+d}) \cap S_{I(x^n)} = \phi\,, \qquad \forall x_{n+1}^{n+d} \in \mathcal{X}_{n+1}^{n+d}$$

- The forbidden points are *concentrated* near the edges of the interval
- “Size” of concentration region depends on the position and length of the encoder’s interval

An Upper Bound on $E_{rd}(P)$

Basic Proof Elements

- Consider *any* interval-mapping encoder
  - Argument extends to arbitrary encoders
- At any time point, how can the encoder map the next *d* symbols?
- Two competing strategies:
  - Short range: *Be faithful to the source* – Likely to generate a large concentration region for the next encoder's interval → *Redundancy*
  - Long range: Map to intervals with a small concentration region – Typically cannot be done while being faithful to the source → *Redundancy*
- Upper bound results from this core tension

Basic Proof Elements – cont.

- The redundancy-delay exponent is upper bounded by

$$E_{rd}(P) \leq 8 \log p_{\min}^{-1}$$

for *almost all* sources P

- $p_{\min} = \min(P)$ is the fastest “zoom-in” rate
- Cannot hold for all sources – For dyadic sources we can attain zero redundancy with zero delay
- Unfortunately, the zero measure set which the result does not cover is larger…
  - For example, includes all binary sources $P = (p, 1-p)$ for which $p = (1+2^k)^{-1}$ for some integer $k \geq 0$
- Convergence to the exponent is not uniform

A Lower Bound on $E_{rd}(P)$

Idealized Arithmetic Coding (AC)

- Interval-mapping encoder with a time-invariant partition
- Relative lengths of subintervals equal symbol probabilities
- Zero asymptotic redundancy
- Some source sequences converge to forbidden points → Maximal delay is *unbounded*
- Analyzing the probability of avoiding all forbidden points in finite time [Shayevitz et al `06]:

$$\mathbb{P}(D(x^n) > d) \leq 4p_{\max}^d(\log p_{\max}^{-1} + O(1))$$

Where $p_{\max} = \max(P)$

- $p_{\max}$ corresponds to the slowest “zoom-in” rate

A Finite Delay AC Variant

- Must intervene in the normal AC process
  - Append 2 fictitious symbols to the source's alphabet
  - Can be mapped to intervals of size ε, so that at least one does not contain a forbidden point
  - The encoder tracks the delay
  - When breeched, inserts the suitable fictitious symbol and the delay is nullified!
- What is the cost in redundancy?
  - *Mismatch* in assigned lengths/probabilities due to fictitious symbols $\log(1-2\varepsilon)^{-1}$
  - Fictitious expected code-length of $\mathbb{P}(D > d)\log\varepsilon^{-1}$
  - Balance by optimizing over ε

A Finite Delay AC Variant – cont.

- The resulting achievable redundancy provides an upper bound on the redundancy-delay function:

$$\mathcal{R}(d, P) \leq c \cdot p_{\max}^{d}(1 + d \log p_{\max}^{-1})^{2}$$

- Thus, a lower bound on the redundancy-delay exponent is given by

$$E_{rd}(P) \geq \log p_{\max}^{-1}$$

Tightening the Lower Bound

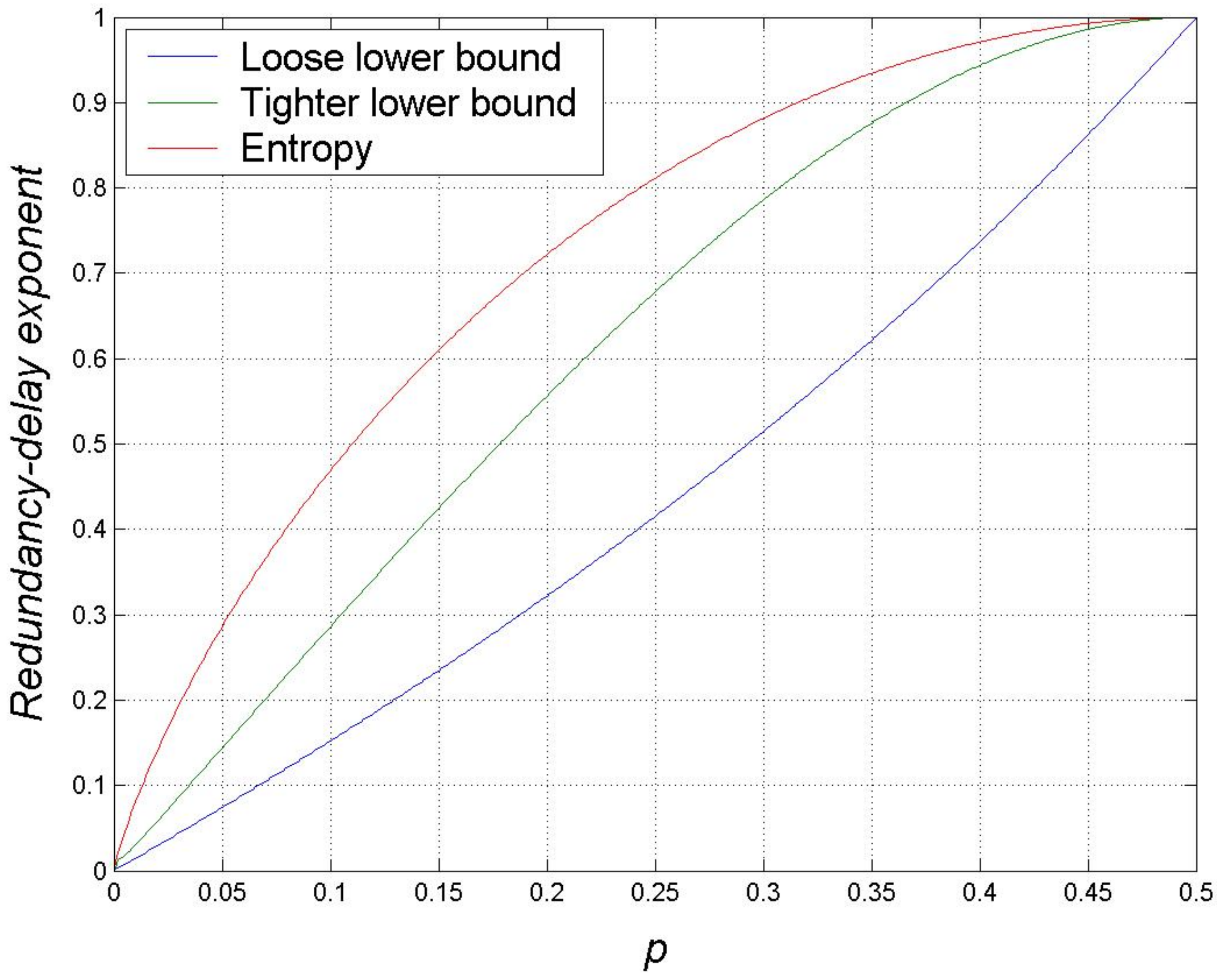
- Via AEP/large deviations

$$E_{rd}(P) \geq \max_{\mu>0} \min \left(H(P) - \mu \,,\ \min_{Q \in A_\mu} 2D(Q\|P) + H(Q) \right)$$

$$A_\mu = \{Q \,:\, D(Q\|P) + H(Q) < H(P) - \mu\}$$

- Requires a *randomized mapping*, to avoid large redundancy under non-typical events

Lower Bounds for Binary Sources



Discussion

- Exponential over polynomial
- Finite horizon n
  - Delay can be much shorter than block/phrase length
  - Best redundancy is $O(n^{-1})$
  - Can meet a delay constraint $d = O(\log n)$ with comparable redundancy
  - Reminiscent observation by [Weinberger *et al* `92]
- Precision vs. redundancy
  - Superior performance over dictionary encoders, at the cost of a *finer precision* for keeping the encoder's state
  - Still, only a *finite precision* is required to obtain the exponential decay!

Further Research

- Upper bound
  - Simplify proof
  - Tighten the bound, eradicate the 8 factor?
  - Extend to all non dyadic sources
- Lower bound
  - Is common randomness necessary for the tightened bound?
  - Can we do even better?