Lossless Data Compression with Side Information: Nonasymptotics and Dispersion

Lampros Gavalakis   Ioannis Kontoyiannis

University of Cambridge

ISIT, 21-26 June 2020

Outline

1. Introduction
   - Reference-based Compression
   - Pair-based Compression
2. Fundamental Limits
3. Coding Theorems for Arbitrary Sources
4. Normal Approximation
   - $R^*(n, \epsilon)$
   - $R^*(n, \epsilon | y_1^n)$
5. Dispersion

Introduction

Compression of the source **X** with side information **Y**:

- *Reference-based* compression
  - Application: Compression of genomic data
    The same reference genome Y is used as side information to compress many source sequences $X^{(1)}, X^{(2)}, \dots$
- *Pair-based* compression
  - Application: Image or video compression
    A new side information sequence Y (previous version/frame) is used every time to compress a new source sequence X

Introduction: Related work

- Generalizations of results from [Kontoyiannis, Verdú, '14]
- Relationship with Slepian-Wolf:
  - Any SW code is a pair-based code
  - Several results in the SW literature, for example [Tan, Kosut, '12], [Jose, Kulkarni, '19], [Chen, Effros, Kostina, '19]
  - Usually random coding in SW
  - Here: deterministic approach, based on the characterization of the optimal compressor with side information

Notation

- X_1^n, Y_1^n : blocks of RVs $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$
- $X_i \sim P_{X_i}$ and $Y_i \sim P_{Y_i}$ take values in $\mathcal{X}$ and $\mathcal{Y}$
- x_1^n, y_1^n : blocks of symbols from $\mathcal{X}^n$ and $\mathcal{Y}^n$
- *Source-side information pair* $(\mathbf{X}, \mathbf{Y}) = \{(X_n, Y_n); n \geq 1\}$ (joint process)

Fundamental Limits

Fixed-to-variable one-to-one compressor with side information of blocklength n:

$$f_n(\cdot|\cdot) : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^*$$

with $f_n(\cdot|y_1^n) : \mathcal{X}^n \to \{0,1\}^*$ one-to-one for each $y_1^n \in \mathcal{Y}^n$

Description length:

$$\ell(f_n(x_1^n|y_1^n)) = \text{length of } f_n(x_1^n|y_1^n) \quad \text{bits}$$

Fundamental Limits

Definition (*Reference-based optimal rate $R^*(n,\epsilon|y_1^n)$*)

$R^*(n,\epsilon|y_1^n)$ is the smallest $R>0$ such that

$$\min_{f_n(\cdot|y_1^n)} \mathbb{P}\left[\ell(f_n(X_1^n|y_1^n)) > nR \,|\, Y_1^n = y_1^n\right] \le \epsilon$$

where the minimum is over all one-to-one compressors $f_n(\cdot|y_1^n)$

Definition (*Pair-based optimal rate $R^*(n,\epsilon)$*)

$R^*(n,\epsilon)$ is the smallest $R>0$ such that

$$\min_{f_n} \mathbb{P}\left[\ell(f_n(X_1^n|Y_1^n)) > nR\right] \le \epsilon$$

where the minimum is over all one-to-one compressors f_n with side information

Fundamental Limits

The optimal compressor f_n^*:

- For each side information string y_1^n, $f_n^*(\cdot|y_1^n)$ is the optimal compressor for $\mathbb{P}(X_1^n = \cdot | Y_1^n = y_1^n)$
- orders the strings x_1^n in order of decreasing probability $\mathbb{P}(X_1^n = x_1^n | Y_1^n = y_1^n)$ and
- assigns to them codewords from $\{0, 1\}^*$ in lexicographic order
- f_n^* achieves the minimum in both definitions

Coding Theorems for Arbitrary Sources

Let $X, Y \sim P_{X,Y}$ be arbitrary discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$

Theorem (*One-shot converse*)

For any compressor with side information f and any integer $k \geq 0$

$$\begin{aligned} &\mathbb{P}\left[\ell(f(X|Y)) \geq k\right] \\ &\geq \sup_{\tau>0} \left\{\mathbb{P}\left[-\log P_{X|Y}(X|Y) \geq k+\tau\right] - 2^{-\tau}\right\} \end{aligned}$$

Coding Theorems for Arbitrary Sources

Theorem (*One-shot achievability*)

There is a compressor f^* such that for all x, y:

$$\ell(f^*(x|y)) \leq -\log P_{X|Y}(x|y)$$

In fact

$$\ell(f^*(x|y)) \leq \log\left(\mathbb{E}\left[\frac{1}{P_{X|Y}(X|y)}\mathbb{I}_{\{P_{X|Y}(X|y)\geq P_{X|Y}(x|y)\}}\middle| Y = y\right]\right)$$

- Both results relate the *description lengths* $\ell(f(x|y))$ to the *conditional information density* $-\log P_{X|Y}(x|y)$

Normal Approximation: Preliminaries

Definition (*Conditional entropy rate*)

$$H(\mathbf{X}|\mathbf{Y}) := \limsup_{n\to\infty} \frac{1}{n} H(X_1^n|Y_1^n) = \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}\left(-\log P(X_1^n|Y_1^n)\right)$$

- If $(\mathbf{X}, \mathbf{Y})$ are jointly stationary, then the above $\limsup$ is in fact a limit

Definition (*Conditional varentropy rate*)

$$\sigma^2(\mathbf{X}|\mathbf{Y}) := \limsup_{n\to\infty} \frac{1}{n} \mathrm{Var}(-\log P(X_1^n|Y_1^n))$$

Lemma

For a broad class of jointly stationary and ergodic source-side information pairs $(\mathbf{X}, \mathbf{Y})$ the above $\limsup$ is in fact the limit

Normal Approximation: $R^*(n,\epsilon)$

Theorem (*Pair-based converse and achievability*)

Suppose $(\mathbf{X},\mathbf{Y})$ is a i.i.d. source-side information pair with $\sigma^2(X|Y) > 0$. For any $0 < \epsilon < \frac{1}{2}$ there are explicit n_1 and $C > 0$ s.t.

$$-\frac{1}{n}C \leq R^*(n,\epsilon) - \left[H(X|Y) + \frac{\sigma(X|Y)}{\sqrt{n}}Q^{-1}(\epsilon) - \frac{\log n}{2n}\right] \leq \frac{1}{n}C$$

for all $n > n_1$

- $H(X|Y) = H(\mathbf{X}|\mathbf{Y})$ is the conditional entropy
- $\sigma^2(X|Y) = \mathrm{Var}\big(-\log P(X|Y)\big) = \sigma^2(\mathbf{X}|\mathbf{Y})$ is the *conditional varentropy*
- n_1 and C depend on the second and third moments of the conditional information density $-\log P_{X|Y}(X|Y)$, the first and second moments of the random variable $\mathrm{Var}[-\log P_{X|Y}(X|Y)|Y]$ and ϵ

Normal Approximation: Reference-based preliminaries

$(\mathbf{X}, \mathbf{Y})$ is a *conditionally-i.i.d. source-side information pair*:
$\mathbf{Y}$ arbitrary and

$$\mathbb{P}(X_1^n = x_1^n | Y_1^n = y_1^n) = \prod_{i=1}^{n} P_{X|Y}(x_i|y_i)$$

Normal Approximation: $R^*(n, \epsilon | y_1^n)$

Theorem (*Reference-based converse and achievability*)

Suppose $(\mathbf{X}, \mathbf{Y})$ is a conditionally-i.i.d. source-side information pair. For any $0 < \epsilon < \frac{1}{2}$ there are explicit $n_0 = n_0(y_1^n)$ and $\zeta_n = \zeta_n(y_1^n) > 0$ s.t.

$$-\frac{1}{n}\zeta_n(y_1^n) \leq R^*(n, \epsilon | y_1^n) - \left[H_n(X|y_1^n) + \frac{\sigma_n(y_1^n)}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} \right] \leq \frac{1}{n}\zeta_n(y_1^n)$$

for all $n > n_0$ and any side information string y_1^n such that $\sigma_n^2(y_1^n) > 0$

- $H_n(X|y_1^n) = \frac{1}{n}\sum_{i=1}^n H(X|Y = y_i)$
- $\sigma_n^2(y_1^n) = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}\big(-\log P(X|y_i)\big| Y = y_i\big)$
- n_0, η and ζ_n depend on the second and third moments of the information densities of the conditional distributions $\{-\log P_{X|Y}(X|y_i)\}_{i=1}^n$ and ϵ

Reference vs Pair-based Rate

First-order term:

In general

$$H_n(X|y_1^n) \neq H(X|Y)$$

although for almost all y_1^∞

$$H_n(X|y_1^n) \to H(X|Y)$$

Second-order term:

Write

$$\hat{H}_X(y) = -\sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \log P_{X|Y}(x|y)$$

and

$$V(y) = \mathrm{Var}[-\log P_{X|Y}(X|y)|Y = y]$$

Reference vs Pair-based Variance

Proposition

$$\sigma^2(X|Y) = \mathbb{E}[V(Y)] + \mathrm{Var}[\hat{H}_X(Y)]$$

In particular, if $(\mathbf{X}, \mathbf{Y})$ is i.i.d. with $(X_n, Y_n) \sim (X, Y)$ and since for almost all y_1^∞

$$\sigma_n^2(y_1^n) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}\big(-\log P(X|y_i)\big| Y = y_i\big) \to \mathbb{E}[V(Y)]$$

we have in general

$$\sigma_n^2(y_1^n) < \sigma^2(X|Y)$$

for typical y's and large n

Normal Approximation: $R^*(n, \epsilon | y_1^n)$

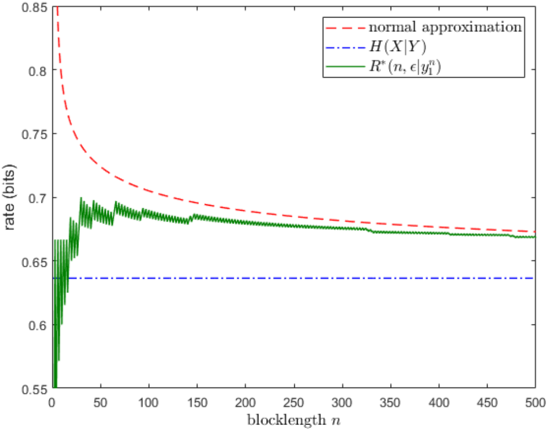

Figure: $\{Y_n\} \sim \text{Bern}(\frac{1}{3})$ i.i.d. $\quad X|Y = 0 \sim \text{Bern}(0.1) \quad X|Y = 1 \sim \text{Bern}(0.6)$
$H(X|Y) \approx 0.636 \qquad H(X) \approx 0.837$
$y_1^n = 001001001 \cdots \qquad \epsilon = 0.1$

Pair-based Dispersion

Definition (*Pair-based Dispersion*)

$$D(\mathbf{X}|\mathbf{Y}) := \limsup_{n\to\infty} \frac{1}{n} \mathrm{Var}\big[\ell(f_n^*(X_1^n|Y_1^n))\big]$$

Theorem

Suppose that both the pair $(\mathbf{X}, \mathbf{Y})$ and $\mathbf{Y}$ itself are irreducible and aperiodic Markov chains, with conditional entropy rate $H(\mathbf{X}|\mathbf{Y})$ and conditional varentropy rate $\sigma^2(\mathbf{X}|\mathbf{Y})$. Then,

$$D(\mathbf{X}|\mathbf{Y}) = \sigma^2(\mathbf{X}|\mathbf{Y})$$

If, moreover, $\sigma^2(\mathbf{X}|\mathbf{Y})$ is nonzero, then:

$$D(\mathbf{X}|\mathbf{Y}) = \lim_{\epsilon\to 0} \lim_{n\to\infty} n \left(\frac{R^*(n,\epsilon) - H(\mathbf{X}|\mathbf{Y})}{Q^{-1}(\epsilon)} \right)^2$$

Reference-based Dispersion

Let $\mathbf{y} = y_1^\infty \in \mathcal{Y}^\infty$

Definition (*Reference-based Dispersion*)

$$D(\mathbf{X}|\mathbf{y}) := \limsup_{n\to\infty} \frac{1}{n}\mathrm{Var}\left[\ell(f_n^*(X_1^n|y_1^n))\right]$$

Theorem

Suppose the side information process $\mathbf{Y}$ is stationary and ergodic, and that the pair $(\mathbf{X}, \mathbf{Y})$ is conditionally i.i.d. Then, for almost any $\mathbf{y}$,

$$D(\mathbf{X}|\mathbf{y}) = \lim_{n\to\infty} \sigma_n^2(y_1^n)$$

If, moreover, $\mathbb{E}[V(Y_1)]$ is nonzero, then, for almost any $\mathbf{y}$:

$$D(\mathbf{X}|\mathbf{y}) = \lim_{\epsilon\to 0}\lim_{n\to\infty} n\left(\frac{R^*(n,\epsilon|y_1^n) - H_n(X|y_1^n)}{Q^{-1}(\epsilon)}\right)^2$$

Conclusions and Further Work

- We gave nonasymptotic normal approximation results in the
    1. Reference-based setting for conditionally i.i.d. source-side information pairs
    2. Pair-based setting for i.i.d. source-side information pairs
- These results remain true if we restrict to prefix-free compressors
- The Pair-based results generalize for Markov source-side information pairs but with a third-order gap
- We gave a characterization of the dispersion in both scenarios
    - For i.i.d. source-side information pairs the reference-based dispersion is in general smaller
    - Does the same hold under more general conditions?
- We have further results, e.g. characterization of the case $\sigma^2(\mathbf{X}|\mathbf{Y}) = 0$ under Markov assumptions
- Further generalizations?
- Is it possible to drop assumptions on the side-infromation process $\mathbf{Y}$?
- More general conditions under which the $\limsup$ is the limit in the definition of the conditional varentropy rate?

Appendix: Reference-based converse

Theorem

For any $0 < \epsilon < \frac{1}{2}$

$$R^*(n, \epsilon | y_1^n) \geq H_n(X | y_1^n) + \frac{\sigma_n(y_1^n)}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - \frac{1}{n}\eta(y_1^n)$$

for all

$$n > \frac{(1 + 6 m_3 \sigma_n^{-3}(y_1^n))^2}{4\big(Q^{-1}(\epsilon)\phi(Q^{-1}(\epsilon))\big)^2}$$

where

$$m_3 = \max_{y \in \mathcal{Y}} \mathbb{E}[| - \log P(X|y) - H(X|y)|^3]$$

and $\eta(y_1^n) = \dfrac{\sigma_n^3(y_1^n) + 6 m_3}{\phi(Q^{-1}(\epsilon))\sigma_n^2(y_1^n)}$

Appendix: Reference-based achievability

Theorem

For any $0 < \epsilon \leq \frac{1}{2}$

$$R^*(n, \epsilon | y_1^n) \leq H_n(X | y_1^n) + \frac{\sigma_n(y_1^n)}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + \frac{1}{n} \zeta_n(y_1^n)$$

for all

$$n > \frac{36 m_3^2}{[\epsilon^2 \sigma_n^6(y_1^n)]}$$

and

$$\zeta_n(y_1^n) = \frac{6 m_3}{\sigma_n^3(y_1^n) \phi\left(\Phi^{-1}\left(\Phi(Q^{-1}(\epsilon)) + \frac{6 m_3}{\sqrt{n}\sigma_n^3(y_1^n)}\right)\right)} + \log\left(\frac{\log e}{\sqrt{2\pi\sigma_n^2(y_1^n)}} + \frac{12 m_3}{\sigma_n^3(y_1^n)}\right)$$

Appendix: Pair-based converse

Theorem

For any $0 < \epsilon < \frac{1}{2}$

$$R^*(n,\epsilon) \geq H(X|Y) + \frac{\sigma(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - \frac{C_1}{n}$$

for all

$$n > \frac{C_1^2}{[4(Q^{-1}(\epsilon))^2 \sigma^2]}$$

where

$$C_1 = \frac{\mathbb{E}[|-\log P(X|Y) - H(X|Y)|^3] + 2\sigma^3}{2\sigma^2 \phi(Q^{-1}(\epsilon))}$$

Appendix: Pair-based achievability

Theorem

For any $0 < \epsilon \leq \frac{1}{2}$

$$R^*(n,\epsilon) \leq H(X|Y) + \frac{\sigma(X|Y)}{\sqrt{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + \frac{C_2}{n}$$

for all

$$n > \frac{4\sigma^2}{B^2 \phi(Q^{-1}(\epsilon))^2} \times \left[\frac{B^2}{2\sqrt{2\pi e}\sigma^2} + \frac{\psi^2}{(1-\frac{1}{2\pi})^2 \bar{v}^2} \right]^2$$

where $\bar{v} = \mathbb{E}[V(Y)]$, $\psi^2 = \mathrm{Var}(V(Y))$, $B = \frac{\mathbb{E}\left[\left|-\log P(X|Y) - H(X|Y)\right|^3\right]}{\sigma^2 \phi(Q^{-1}(\epsilon))}$ *and*

$$C_2 = \log\left(\frac{2}{\bar{v}^{1/2}} + \frac{24 m_3 (2\pi)^{3/2}}{\bar{v}^{3/2}} \right) + B$$