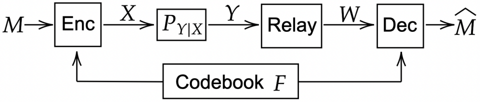
Nonasymptotic Oblivious Relaying and Variable-Length Noisy Lossy Source Coding

Yanxiao Liu*, Sepehr Heidari Advary† and Cheuk Ting Li*

* The Chinese University of Hong Kong
† Sharif University of Technology

Background

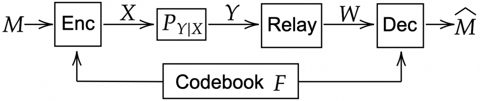


The oblivious relay channel (a.k.a. information bottleneck (IB) channel) concerns a channel coding setting where

- the decoder does not directly observe the channel output;
- rather, the channel output is relayed to the decoder via a rate-limited link
- by an <u>oblivious</u> relay, which does not know the codebook.

Applications: cloud radio access networks where base stations are connected to a cloud-computing central processor via error-free rate-limited fronthaul links.

Background



- M in encoded into $X^n = (X_1, \ldots, X_n)$ and sent through memoryless $P_{Y|X}$.
- An oblivious relay (no codebook) receives Y^n and sends description W.
- The decoder observes W and decodes M.

Theorem (Sanderovich et al., 2008)

The minimum asymptotic description rate (as the blocklength $n \to \infty$) needed to support a message transmission rate C is given by the information bottleneck

$$\mathrm{IB}_{X\to Y}(\mathsf{C}) := \min_{P_{U|Y}:\, I(X;U)\geq \mathsf{C}} I(Y;U), \qquad X \leftrightarrow Y \leftrightarrow U.$$

Our Contributions

- We provide novel nonasymptotic achievability results of:
  1. variable-length noisy lossy source coding;
  2. the information bottleneck (IB) channel where the relay communicates to the decoder via fixed-length or variable-length codes.

Overview: Finite-Blocklength Results

We study the trade-off between description rate, message rate and blocklength n.

For the IB channel with fixed-length description and error probability is ϵ, a rate

$$\mathrm{IB}(\mathsf{C}) + \sqrt{\frac{1}{n}\mathrm{VIB}(\mathsf{C})}Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right) \tag{1}$$

suffices, where $\mathrm{VIB}_{X\to Y}(\mathsf{C}) := \mathrm{Var}\left[\iota_{Y;U}(Y;U) - \lambda^* \iota_{X;U}(X;U)\right]$ is a second-order version of the information bottleneck, $\iota_{Y;U}(y;u) := \log\left(P_{U|Y}(u|y)/P_U(u)\right)$ is computed by the optimal $P_{U|Y}$ in $\mathrm{IB}_{X\to Y}(\mathsf{C})$, and $\lambda^* := \frac{\mathrm{d}}{\mathrm{d}\mathsf{C}}\mathrm{IB}_{X\to Y}(\mathsf{C})$.

For the IB channel with variable-length description, it suffices to use a rate

$$(1-\epsilon)\left(\mathrm{IB}(\mathsf{C}) + \sqrt{\frac{\ln n}{n}\mathrm{CVIB}(\mathsf{C})}\right) + O\left(\frac{1}{\sqrt{n}}\right) \tag{2}$$

where $\mathrm{CVIB}_{X\to Y}(\mathsf{C}) := \mathbb{E}\left[\mathrm{Var}\left[\lambda^* \iota_{X;U}(X;U) \,\middle|\, Y, U\right]\right]$ is *conditional-var-IB*.

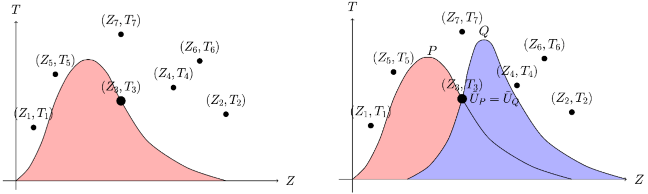
Comparing (1) and (2), we see the fixed- and variable-length cases differ vastly.

Poisson Functional Representation and Poisson Matching Lemma

Poisson Functional Representation (PFR) (Li and El Gamal (2018))

- Let $(T_i)_i$ be a Poisson process with rate 1, i.e., $T_1, T_2 - T_1, \ldots \sim \mathrm{Exp}(1)$.
- Let $(\bar{Z}_i)_i \overset{iid}{\sim} P_{\bar{Z}}$ be a sequence independent of $(T_i)_i$.
- Fix distribution P over $\mathcal{U}$ s.t. $P \ll P_{\bar{Z}}$ (P is absolutely continuous w.r.t $P_{\bar{Z}}$).

$$\tilde{U}_P := \bar{Z}_K, \quad \text{where} \quad K = \operatorname{argmin}_i \Big(T_i \cdot \big(\frac{\mathrm{d}P}{\mathrm{d}P_{\bar{Z}}}(\bar{Z}_i)\big)^{-1} \Big).$$



Poisson Matching Lemma (Li and Anantharam, 2021)

Consider two distributions $P, Q \ll P_{\bar{Z}}$, almost surely, we have

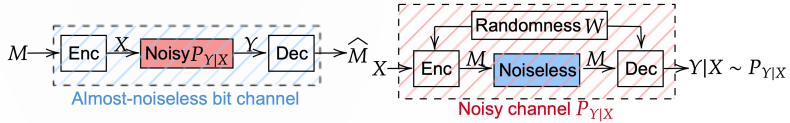
$$\mathbb{P}\big(\tilde{U}_Q \neq \tilde{U}_P \,|\, \tilde{U}_P\big) \leq 1 - \left(1 + \frac{\mathrm{d}P}{\mathrm{d}Q}(\tilde{U}_P)\right)^{-1}.$$

Poisson Functional Representation and Channel Simulation

The "marked" PP has a "query operation": with input P, it gives $\tilde{U}_P \sim P$.

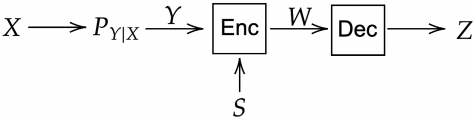
One-Shot Channel Simulation

- Alice sees $X \sim P_X$ and sends description M to Bob noiselessly, so that Bob generates $Y \sim P_{Y|X}$. They share unlimited common randomness.
- The goal is to find the minimum expected description length of M.
- **Achievability**: By PFR, $H(K) \leq I(X;Y) + \log\big(I(X;Y)+1\big) + 4$.



Noisy Lossy Source Coding

- We first study the one-shot noisy lossy source coding.
- We will utilize the result to analyze the IB channel.



- For $(X, Y) \sim P_{X,Y}$, encoder observes Y and sends a description $W = f(S, Y)$, where $S \sim P_S$ is local randomness independent of Y.
- Decoder observes W and recovers $Z = g(W)$.
- W can be fixed-length or variable-length!
- The goal is to have a small $P_e := \mathbb{P}(d(X, Z) > \mathsf{D})$, where $d : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ is the distortion measure and $\mathsf{D} \in \mathbb{R}$.

Noisy Lossy Source Coding

For fixed-length case $W \in [\mathsf{L}]$, it has been studied by Kostina and Verdú (2016).

Theorem 1 (Kostina and Verdú (2016))

For any $P_{\bar{Z}}$ and $\gamma > 0$, there exists a fixed-length code with

$$P_e \leq \mathbb{P}\Big(\psi_{\bar{Z}}(Y, \mathsf{D}, T) \geq \log \gamma\Big) + e^{-\mathsf{L}/\gamma},$$

where $T \sim \mathrm{Unif}(0,1)$ is independent of Y, $Z \perp\!\!\!\perp (X, Y)$ and

$$\psi_{\bar{Z}}(y, \mathsf{D}, t) := \inf_{P_Z :\, \mathbb{P}(d(X,Z) > \mathsf{D} | Z, Y=y) \leq t \text{ a.s.}} D(P_Z \| P_{\bar{Z}}). \tag{3}$$

We study the variable-length case where $W \in \mathcal{C}$ lies in a prefix-free codebook $\mathcal{C} \subseteq \{0,1\}^* := \cup_{k=0}^{\infty} \{0,1\}^k$; the goal is to minimize the expected length $\mathbb{E}[|W|]$.

Theorem 2

For any $P_{\bar{Z}}$, $\epsilon' > 0$, and function $\beta : \mathcal{Y} \to [0,1]$, there is a variable-length code with $P_e \leq \mathbb{E}[\beta(Y)] + \epsilon'$ and $\ell(t) := t + \log(t+2) + 4$,

$$\mathbb{E}[|W|] \leq \ell\left(\mathbb{E}\Big[(1-\beta(Y))\psi_{\bar{Z}}(Y, \mathsf{D}, \epsilon')\Big]\right),$$

assuming the expectation above is finite, and $\psi_{\bar{Z}}(\cdot)$ is defined in (3).

Noisy Lossy Source Coding: Proof

Theorem 2

For any $P_{\bar{Z}}$, $\epsilon' > 0$, and function $\beta : \mathcal{Y} \to [0,1]$, there is a variable-length code with $P_e \leq \mathbb{E}[\beta(Y)] + \epsilon'$ and $\ell(t) := t + \log(t+2) + 4$,

$$\mathbb{E}[|W|] \leq \ell\left(\mathbb{E}\left[(1-\beta(Y))\psi_{\bar{Z}}(Y, \mathsf{D}, \epsilon')\right]\right).$$

- Consider channel $P_{\hat{Z}|Y}$, which conditional on $Y = y$, $\hat{Z}$ has same distribution as $\bar{Z} \sim P_{\bar{Z}}$ conditional on $\phi(y, \bar{Z}, \mathsf{D}) := \mathbb{P}(d(X, z) > \mathsf{D} | Y = y) \leq \epsilon'$.
- Let PFR take $K := \mathrm{argmin}_k T_k \big/ \left(P_{\hat{Z}|Y}(\bar{Z}_k|Y)/P_{\bar{Z}}(\bar{Z}_k)\right)$. Construct:
  - Encoder observes Y, outputs $\tilde{K} = K$ w.p. $1 - \beta(Y)$, or $\tilde{K} = 1$ w.p. $\beta(Y)$,
  - then encodes $\tilde{K}$ by an optimal prefix-free code into W.
  - Decoder decodes $\tilde{K}$ from W, outputs $Z = \bar{Z}_K$.
- (Li, 2024): $\bar{Z}_K | \{Y = y\} \sim P_{\hat{Z}|Y}(\cdot|y)$ and $\mathbb{E}[\log K | Y = y] \leq \psi_{\bar{Z}}(y, \mathsf{D}, \epsilon') + 1$,

$$\mathbb{P}(d(X,Z) > \mathsf{D}) \leq \mathbb{E}[\beta(Y)] + \mathbb{E}[\phi(Y, \bar{Z}_K, \mathsf{D})] \leq \mathbb{E}[\beta(Y)] + \epsilon',$$

$$\mathbb{E}[\log \tilde{K}] \leq \mathbb{E}\left[(1-\beta(Y))\psi_{\bar{Z}}(Y, \mathsf{D}, \epsilon')\right] + 1.$$

- Max-entropy dist of $\mathbb{E}[\log \tilde{K}]$ gives $H(\tilde{K}) \leq \ell(\mathbb{E}[(1-\beta(Y))\psi_{\bar{Z}}(y, \mathsf{D}, \epsilon')]) - 2$.
- Huffman code gives $\mathbb{E}[|W|] \leq H(\tilde{K}) + 1$; resolving common randomness $(\bar{Z}_i, T_i)_i$ also incurs 1-bit penalty on $\mathbb{E}[|W|]$ (Li and El Gamal, 2018).

Noisy Lossy Source Coding: Block Setting

When $X = X^n, Y = Y^n, Z = Z^n$ are sequences, $(X_i, Y_i) \overset{iid}{\sim} P_{X,Y}$ and $d(x^n, z^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, z_i)$. For fixed-length, as $n \to \infty$, the optimal description rate is

$$R(\mathsf{D}) := \min_{P_{Z|Y}:\, \mathbb{E}[d(X,Z)] \leq \mathsf{D}} I(Y;Z), \tag{4}$$

by Dobrushin and Tsybakov (1962). Let $\lambda^* := -R'(\mathsf{D})$ and $P_{Z^*|Y}$ attain its min.

Theorem 3: Fixed Length Coding (Kostina and Verdú (2016))

The smallest L s.t. there exists a fixed-length code with blocklength n, $P_e \leq \epsilon$ is

$$nR(\mathsf{D}) + \sqrt{n\tilde{\mathrm{V}}(\mathsf{D})}Q^{-1}(\epsilon) + O(\log n),$$

where $\tilde{\mathrm{V}}(\mathsf{D}) := \mathrm{Var}[\iota_{Y;Z^*}(Y;Z^*) + \lambda^* d(X, Z^*)]$.

Theorem 4: Variable Length Coding

For $\epsilon > 0$, if $n \geq n_0$, there exists a variable-length code with $P_e \leq \epsilon$, and

$$\mathbb{E}[|W|] \leq (1-\epsilon)\left(nR(\mathsf{D}) + \sqrt{(n \ln n)\widetilde{\mathrm{CV}}(\mathsf{D})}\right) + O(\sqrt{n}),$$

where $\widetilde{\mathrm{CV}}(\mathsf{D}) := (\lambda^*)^2\mathbb{E}[\mathrm{Var}[d(X, Z^*) \,|\, Y, Z^*]]$.

Assume above theorems are subject to the regularity conditions in (Kostina and Verdú, 2016).
In Theorem 4, n_0 and the constant in $O(\sqrt{n})$ depends on $P_{X,Y}, d, \mathsf{D}, \epsilon$

Noisy Lossy Source Coding: Block Setting

Theorem 4: Variable Length Coding

For $\epsilon > 0$, if $n \geq n_0$, there exists a variable-length code with $P_e \leq \epsilon$, and

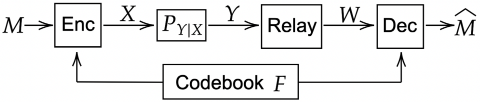
$$\mathbb{E}[|W|] \leq (1-\epsilon)\left(nR(\mathsf{D}) + \sqrt{(n \ln n)\widetilde{\mathrm{CV}}(\mathsf{D})}\right) + O(\sqrt{n}),$$

where $\widetilde{\mathrm{CV}}(\mathsf{D}) := (\lambda^*)^2 \mathbb{E}[\mathrm{Var}[d(X, Z^*) \mid Y, Z^*]]$.

The proof is based on (Kostina and Verdú, 2016) and Theorem 2.

- Theorem 4 exhibits a different behavior compared to Theorem 3:
  - The asymptotic rate is $(1-\epsilon)R(\mathsf{D})$ instead of $R(\mathsf{D})$,
  - similar to the phenomenon in lossless/lossy compression with error (Koga and Yamamoto, 2005; Kostina et al., 2015).
- Intuitively, we discard a portion ϵ of Y^n by assigning the same short codeword, inducing error probability ϵ while reducing expected length $\approx \epsilon R(\mathsf{D})$.

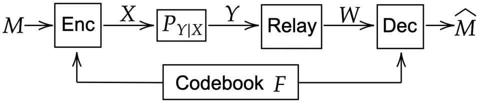
Information Bottleneck Channel



One-shot information bottleneck channel:

- Encoder and decoder (but not relay) share random i.i.d. codebook $F \sim P_F$.
- Encoder observes $M \sim \mathrm{Unif}([\mathsf{L}])$; sends $X = f(F, M)$ to a channel $P_{Y|X}$.
- Relay observes Y and sends description $W = f_r(S, Y)$ noiselessly, $S \sim P_S$.
- Decoder observes F and W, and recovers $\hat{M} = g(F, W) \in [\mathsf{L}]$.
- The goal is to minimize the error probability $P_e := \mathbb{P}(M \neq \hat{M})$.

Information Bottleneck Channel



For description W, we study two settings:

- Fixed-length: $W \in [\mathsf{K}]$, and we want to minimize $\mathsf{K} \in \mathbb{N}$;
- Variable-length: W is in codebook $\mathcal{C}_W \subseteq \{0,1\}^*$, and we minimize $\mathbb{E}[|W|]$.

We also study **block case**: encoder sends X^n to memoryless $P_{Y|X}^n$ and relay sees Y^n (i.e., substitute $X = X^n$, $Y = Y^n$ and $P_{Y|X} = P_{Y|X}^n$ in the one-shot case).

- Define $\mathsf{B}_{\mathrm{F}}^*(n, \mathsf{C}, \epsilon)$ be the smallest possible relay description rate $n^{-1} \log \mathsf{K}$ among fixed-length schemes with message rate $n^{-1} \log \mathsf{L} \geq \mathsf{C}$.
- Define $\mathsf{B}_{\mathrm{V}}^*(n, \mathsf{C}, \epsilon)$ for variable-length schemes similarly.

When $n \to \infty$, the asymptotic capacity for fixed-length case is $\mathrm{IB}_{X \to Y}(\mathsf{C})$:

$$\limsup_{\epsilon \to 0} \limsup_{n \to \infty} \mathsf{B}_{\mathrm{F}}^*(n, \mathsf{C}, \epsilon) = \min_{P_{U|Y} : I(X;U) \geq \mathsf{C}} I(Y;U), \qquad X \leftrightarrow Y \leftrightarrow U.$$

We will utilize the noisy lossy source coding for nonasymptotic results of IB channel.

Information Bottleneck Channel: Variable-length Description

For the one-shot variable-length setting: if $P_{U|Y}$ achieves the optimum of IB, let:

1. Relay performs one-shot channel simulation to simulate $P_{U|Y}$,
   - expected description length bounded by $I(Y;U) + \log(I(Y;U)+2) + 3$.
2. Decoder performs channel decoding for $P_{U|X}$ to recover U.

Theorem 5

Fix any P_X, $P_{U|Y}$ and $\epsilon' \geq 0$. There is a one-shot variable-length scheme with

$$P_e \leq \mathbb{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;U}(X;U)}, 1\right\}\right)^{(\mathsf{L}+1)/2}\right] + \epsilon', \quad (5)$$

and $\mathbb{E}[|W|] \leq \ell((1-\epsilon')I(Y;U))$, where $\ell(t) := t + \log(t+2) + 4$.

Theorem 5 on asymptotic setting $X = X^n$, $Y = Y^n$ yields the asymptotic result

$$\limsup_{n\to\infty} \mathsf{B}^*_{\mathrm{V}}(n, \mathsf{C}, \epsilon) \leq (1-\epsilon)\mathrm{IB}_{X\to Y}(\mathsf{C}).$$

To prove Theorem 5, we use one-shot channel simulation, one-shot channel coding, and the strategy to resolve the common randomness in (Li and El Gamal, 2018).

Information Bottleneck Channel: Variable-length Description

Theorem 5

Fix any P_X, $P_{U|Y}$ and $\epsilon' \geq 0$. There is a one-shot variable-length scheme with

$$P_e \leq \mathbb{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;U}(X;U)}, 1\right\}\right)^{(\mathsf{L}+1)/2}\right] + \epsilon', \tag{6}$$

and $\mathbb{E}[|W|] \leq \ell((1-\epsilon')I(Y;U))$, where $\ell(t) := t + \log(t+2) + 4$.

- Channel simulation on $P_{U|Y}$ gives $K = f_a(S_a, Y)$ for encoding and $U = g_a(S_a, K)$ for decoding. U follows $P_{U|Y}$ given Y, and $\mathbb{E}[\log K] \leq I(Y;U)+1$.
- Let $\tilde{K} = K$ w.p. $1-\epsilon'$ and $\tilde{K} = 1$ w.p. ϵ'; $\mathbb{E}[|W|] \leq \ell((1-\epsilon')I(Y;U)) - 1$.
- Channel coding on $P_{U|X}$ gives $X = f_b(F, M)$ for encoding and $\hat{M} = g_b(F, U)$ for decoding; $\mathbb{P}(M \neq \hat{M}) \leq \mathbb{E}[1 - (1 - \min\{2^{-\iota_{X;U}(X;U)}, 1\})^{(\mathsf{L}+1)/2}]$.
- Our scheme for the IB channel:
  1. Encoder produces $X = f_b(F, M)$;
  2. Relay computes $K = f_a(S_a, Y)$, generates $\tilde{K}$ and encodes it to W;
  3. Decoder recovers $\tilde{K}$, $U = g_a(S_a, \tilde{K})$, $\hat{M} = g_b(F, U)$.
- Since $\mathbb{P}(K \neq \tilde{K}) \leq \epsilon'$, using $\tilde{K}$ instead of K increases P_e by at most ϵ'.
- Resolving common randomness incurs 1-bit penalty on $\mathbb{E}[|W|]$.

Information Bottleneck Channel via Noisy Lossy Source Coding

We can further refine the bound by utilizing the noisy lossy source coding (NLSC)!

- For any fixed $p_{U|Y}$ the relay performs a noisy lossy source coding on Y;
- Decoder recovers $\hat{U}$, with a distortion function $d(x,\hat{u}) = -\iota_{X;U}(x;\hat{u})$.

Theorem 6

Fix any P_X, $P_{U|Y}$, $\mathsf{C}, \epsilon' > 0$, and function $\beta : \mathcal{Y} \to [0,1]$. There is a one-shot variable-length scheme with message size L,

$$P_e \leq \mathbb{E}[\beta(Y)] + 2^{-\mathsf{C}}(\mathsf{L}+1)/2 + \epsilon',$$
$$\mathbb{E}[|W|] \leq \ell\Big(\mathbb{E}\Big[(1-\beta(Y))\psi_U(Y,\mathsf{C},\epsilon')\Big]\Big),$$

assuming the expectation above is finite, where $\ell(t) := t + \log(t+2) + 4$, and

$$\psi_U(y,\mathsf{C},t) := \inf_{P_{\tilde{U}}:\, \mathbb{P}(\iota_{X;U}(X;\tilde{U})<\mathsf{C}|\tilde{U},Y=y)\leq t \text{ a.s.}} D(P_{\tilde{U}} \| P_U). \tag{7}$$

(We assume $\tilde{U} \perp\!\!\!\perp (X,Y)$ above.)

Information Bottleneck Channel via Noisy Lossy Source Coding

Theorem 6

Fix any P_X, $P_{U|Y}$, $\mathsf{C}, \epsilon' > 0$, and function $\beta : \mathcal{Y} \to [0,1]$. There is a one-shot variable-length scheme with message size L and $\ell(t) := t + \log(t+2) + 4$,

$$P_e \leq \mathbb{E}[\beta(Y)] + 2^{-\mathsf{C}}(\mathsf{L}+1)/2 + \epsilon',$$

$$\mathbb{E}[|W|] \leq \ell\Big(\mathbb{E}\Big[(1-\beta(Y))\psi_U(Y, \mathsf{C}, \epsilon')\Big]\Big).$$

- Fix any P_X, $P_{U|Y}$. Apply Theorem 2 on $d(x, \hat{u}) = -\iota_{X;U}(x; \hat{u})$, $\mathsf{D} = -\mathsf{C}$.
- Relay has $W = f_r(S, Y)$ for encoding, decoder has $\hat{U} = g_r(W)$ for decoding,

$$\mathbb{P}(d(X; \hat{U}) > \mathsf{D}) = \mathbb{P}(\iota_{X;U}(X; \hat{U}) < \mathsf{C}) \leq \mathbb{E}[\beta(Y)] + \epsilon', \tag{8}$$

$$\mathbb{E}[|W|] \leq \ell(\mathbb{E}[(1-\beta(Y))\psi_U(Y, \mathsf{C}, \epsilon')]).$$

- We use PFR to design encoder & decoder: let $(\bar{X}_i)_i \overset{iid}{\sim} P_X$, $(T_i)_i \sim \mathrm{PP}(1)$:
  1. Encoder observes $M \in [\mathsf{L}]$ and sends $X = \bar{X}_M$.
  2. Decoder sees $\hat{U} = g_r(W)$ and gets $\hat{M} = \mathrm{argmin}_{k \in [\mathsf{L}]} \frac{T_k}{P_{X|U}(\bar{X}_k | \hat{U}) / P_X(\bar{X}_k)}$.
- By Poisson matching lemma (Li and Anantharam, 2021) and (8),

$$\mathbb{P}(M \neq \hat{M} | M = m) \leq \mathbb{E}[\beta(Y)] + 2^{-\mathsf{C}}(\mathsf{L}+1)/2 + \epsilon'.$$

Information Bottleneck Channel via Noisy Lossy Source Coding

We then study the block setting, still utilizing the noisy lossy source coding.

Theorem 7

Fix any P_X, $\epsilon > 0$ and $0 < \mathsf{C} < I(X;Y)$. Under the regularity conditions in the footnote,[a] we have

$$\mathsf{B}^*_{\mathrm{V}}(n, \mathsf{C}, \epsilon) \leq (1-\epsilon)\left(\mathrm{IB}(\mathsf{C}) + \sqrt{\frac{\ln n}{n}\mathrm{CVIB}(\mathsf{C})}\right) + O\left(\frac{1}{\sqrt{n}}\right),$$

where $\mathrm{IB}(\mathsf{C}) = \mathrm{IB}_{X\to Y}(\mathsf{C})$ and $\mathrm{CVIB}(\mathsf{C}) = \mathrm{CVIB}_{X\to Y}(\mathsf{C})$ have been defined:

$$\mathrm{IB}_{X\to Y}(\mathsf{C}) := \min_{P_{U|Y}:\, I(X;U)\geq \mathsf{C}} I(Y;U)$$

$$\mathrm{CVIB}_{X\to Y}(\mathsf{C}) := \mathbb{E}\Big[\mathrm{Var}\Big[\lambda^* \iota_{X;U}(X;U)\,\Big|\, Y, U\Big]\Big].$$

[a] We need $\tilde{R}(\mathsf{C}) := \min_{P_{\tilde{U}|Y}:\, \mathbb{E}[\iota_{X;U}(X,\tilde{U})]\geq \mathsf{C}} I(Y;\tilde{U})$ to be twice continuously differentiable as a function of P_Y (assuming $P_{X,Y} = P_{X|Y}P_Y$, and let $P_{U|Y}$ be the minimizer in $\mathrm{IB}(\mathsf{C})$), and perturbing P_Y within a neighborhood of the original P_Y will not affect the support of U^*, where $P_{U^*|Y}$ attains the minimum in $\tilde{R}(\mathsf{C})$.

Information Bottleneck Channel via Noisy Lossy Source Coding

Theorem 7

Fix any P_X, $\epsilon > 0$ and $0 < \mathsf{C} < I(X;Y)$. Under the regularity conditions,

$$\mathsf{B}^*_{\mathrm{V}}(n, \mathsf{C}, \epsilon) \leq (1-\epsilon)\left(\mathrm{IB}(\mathsf{C}) + \sqrt{\frac{\ln n}{n}\mathrm{CVIB}(\mathsf{C})}\right) + O\left(\frac{1}{\sqrt{n}}\right),$$

where $\mathrm{IB}(\mathsf{C}) = \mathrm{IB}_{X\to Y}(\mathsf{C})$ and $\mathrm{CVIB}(\mathsf{C}) = \mathrm{CVIB}_{X\to Y}(\mathsf{C})$.

- Let $P_{U|Y}$ achieve the optimum in $\mathrm{IB}(\mathsf{C})$, define $d(x,u) = -\iota_{X;U}(x;u)$.
- Consider $R(\mathsf{D}) = \min_{P_{\tilde{U}|Y}:\,\mathbb{E}[d(X,\tilde{U})]\leq \mathsf{D}} I(Y;\tilde{U})$ of NLSC at $\mathsf{D} = -\mathsf{C}$.
- We can check $I(X;\tilde{U}) \geq \mathbb{E}[\iota_{X;U}(X,\tilde{U})]$, implying $P_{U|Y}$ achieves the optimum in $R(\mathsf{D})$ and $-R'(\mathsf{D}) = \mathrm{IB}'(\mathsf{C}) =: \lambda^*$.
- By Theorem 4 (block setting of NLSC),

$$\mathbb{E}[|W|] \leq (1-\epsilon)\left(n\mathrm{IB}(\mathsf{C} + \sqrt{(n\ln n)\mathrm{CVIB}(\mathsf{C})}\right) + O(\sqrt{n}),$$

  with decoded $\hat{U}^n$ satisfying $\mathbb{P}\left(\iota(X^n;\hat{U}^n) < n\mathsf{C} + \log n\right) \leq \epsilon - 1/\sqrt{n}$.
- By Poisson matching lemma (like Theorem 6) on $X^n, \hat{U}^n$ and $\mathsf{L} = \lceil 2^{n\mathsf{C}} \rceil$,

$$\mathbb{P}(M \neq \hat{M}|M = m) \leq 2^{-(n\mathsf{C}+\log n)}m + \epsilon - 1/\sqrt{n} \leq \epsilon.$$

Information Bottleneck Channel: Fixed-length Description

We now consider the fixed-length case where $W \in [\mathsf{K}]$.
Similar to Theorem 6 (IB channel via NLSC), we have the following theorem.

Theorem 8

Fix P_X, $P_{U|Y}$ and $\mathsf{C}, \gamma > 0$. There is a one-shot fixed-length scheme with message size L, description size K,

$$P_e \leq \mathbb{P}\Big(\psi_U(Y, \mathsf{C}, T) \geq \log \gamma\Big) + 2^{-\mathsf{C}}(\mathsf{L}+1)/2 + e^{-\mathsf{K}/\gamma},$$

where $T \sim \mathrm{Unif}(0,1)$, $T \perp\!\!\!\perp Y$,and

$$\psi_U(y, \mathsf{C}, t) := \inf_{P_{\tilde{U}}:\, \mathbb{P}(\iota_{X;U}(X;\tilde{U}) < \mathsf{C} | \tilde{U}, Y=y) \leq t \text{ a.s.}} D(P_{\tilde{U}} \| P_U).$$

Information Bottleneck Channel: Fixed-length Description

We can also obtain a second-order result in terms of $\mathrm{VIB}(\mathsf{C})$.

Theorem 9

Fix any P_X, $\epsilon > 0$ and $0 < \mathsf{C} < I(X;Y)$. Under some regularity conditions, we have

$$\mathsf{B}^*_{\mathrm{F}}(n, \mathsf{C}, \epsilon) \leq \mathrm{IB}(\mathsf{C}) + \sqrt{\frac{1}{n}\mathrm{VIB}(\mathsf{C})}Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right),$$

where $\mathrm{IB}(\mathsf{C}) = \mathrm{IB}_{X\to Y}(\mathsf{C})$ and $\mathrm{VIB}(\mathsf{C}) = \mathrm{VIB}_{X\to Y}(\mathsf{C})$.

Information Bottleneck Channel: Fixed-length Description

Summary

- We have provided novel results for variable-length noisy lossy source coding.
- We have derived nonasymptotic achievability results for the IB channel
  1. with both fixed and variable-length description cases,
  2. using techniques in noisy lossy source coding and Poisson functional representation.

References

Dobrushin, R. and Tsybakov, B. (1962). Information transmission with additional noise. *IRE Transactions on Information Theory*, 8(5):293–304.

Koga, H. and Yamamoto, H. (2005). Asymptotic properties on codeword lengths of an optimal FV code for general sources. *IEEE Transactions on Information Theory*, 51(4):1546–1555.

Kostina, V., Polyanskiy, Y., and Verdú, S. (2015). Variable-length compression allowing errors. *IEEE Transactions on Information Theory*, 61(8):4316–4330.

Kostina, V. and Verdú, S. (2016). Nonasymptotic noisy lossy source coding. *IEEE Transactions on Information Theory*, 62(11):6111–6123.

Li, C. T. (2024). Pointwise redundancy in one-shot lossy compression via Poisson functional representation. In *arXiv preprint; short version presented at 2024 International Zurich Seminar on Information and Communication*, pages 28–29.

Li, C. T. and Anantharam, V. (2021). A unified framework for one-shot achievability via the Poisson matching lemma. *IEEE Transactions on Information Theory*, 67(5):2624–2651.

Li, C. T. and El Gamal, A. (2018). Strong functional representation lemma and applications to coding theorems. *IEEE Transactions on Information Theory*, 64(11):6967–6978.