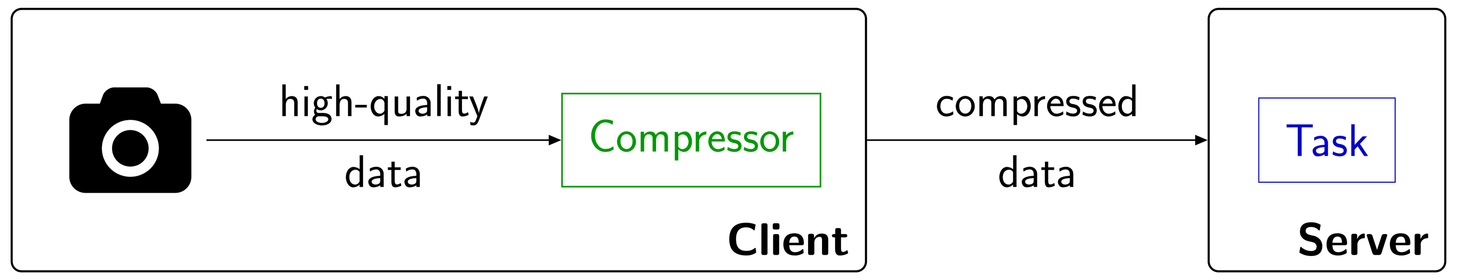
Single-Shot Compression for Hypothesis Testing

NYU TANDON SCHOOL OF ENGINEERING

Fabrizio Carpi, Siddharth Garg, Elza Erkip

NYU WIRELESS

Introduction



Motivation: a resource constrained *client* offloads costly task-related computations to remote a *server* (edge/cloud computing).

Open need: design task-aware source coding schemes which provides *effective* representations of the source data.

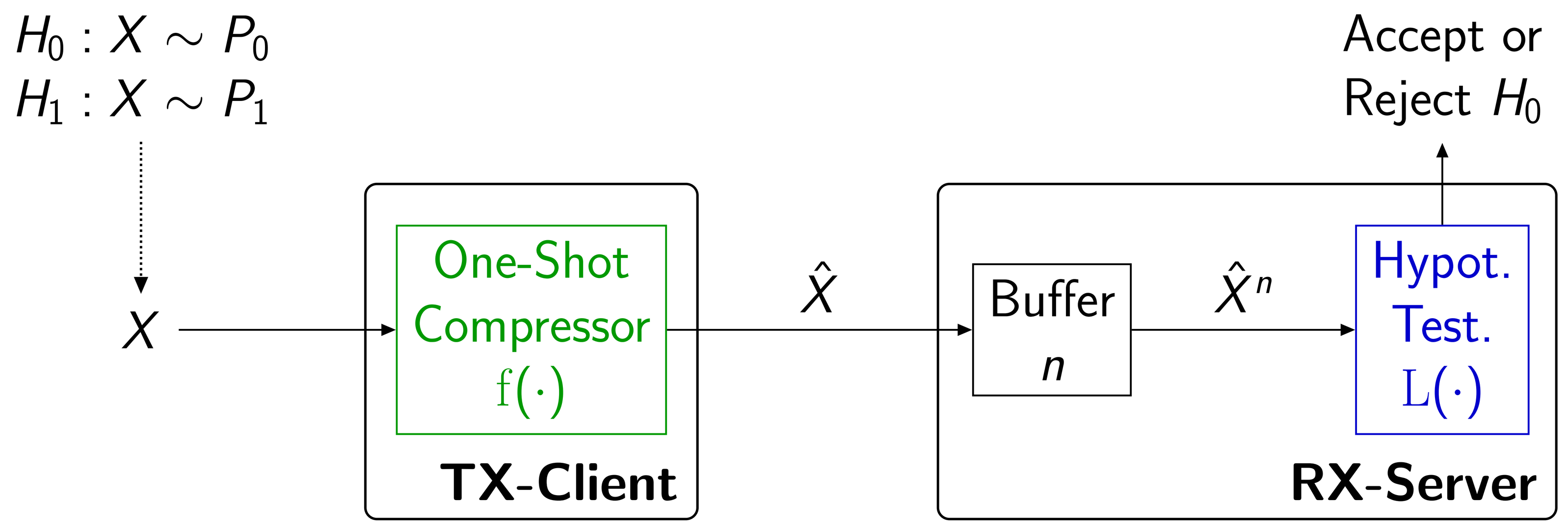
Assumptions:

- task: binary hypothesis testing;
- client: constrained device which cannot perform task locally, does not have memory and can only do simple scalar compression;
- server: hypothesis testing on a block of compressed samples.

Our work: single-shot fixed-length compression for hypothesis testing.

- problem formulation;
- analyze the error performance;
- propose a task-oriented compression algorithm for hypothesis testing.

System Model



Source	Compressor	Hypothesis Testing
$x \in \mathcal{X} = \{1, \dots, \lvert\mathcal{X}\rvert\}$ $X \sim P_\theta(x)$, $\theta \in \{0,1\}$	$\mathrm{f} : \mathcal{X} \to \mathcal{M} = \{1, \dots, M\}$ $\hat{X} = \mathrm{f}(X)$, $\hat{X} \sim \hat{P}_\theta(\hat{X})$	$\mathrm{L}(\hat{X}^n) \overset{\hat{\theta}=0}{\underset{\hat{\theta}=1}{\gtrless}} \log T$

Fixed rate compression $R = \log M$. We consider $M < \lvert\mathcal{X}\rvert$.

Task: binary hypothesis testing.

- if type-I error $< \epsilon$, then type-II error β_n^ϵ (accept H_0 when H_1 is true) decays exponentially in n as $\gamma = -\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\epsilon$;
- **our performance metric:** type-II error exponent γ;
- Chernoff-Stein [1]: optimal type-II error exponent is $\gamma^\star = D(P_0||P_1)$ when there is no compression;
- with compression: error exponent depends on (f, R): $\gamma_\mathrm{f}(R)$;
- **compression penalty:** $\Delta_\mathrm{f}(R) = D(P_0||P_1) - \gamma_\mathrm{f}(R)$.

Hypothesis Testing under Single-shot Compression

Hypothesis test on $\hat{X} \sim \hat{P}_\theta$:

- log-likelihood ratio test on $\hat{X}^n$ is optimal;
- optimal error exponent is $\gamma_\mathrm{f}^\star(R) = D(\hat{P}_0||\hat{P}_1)$.

$\Longrightarrow$ **Compression penalty:** $\Delta_\mathrm{f}(R) = D(P_0||P_1) - D(\hat{P}_0||\hat{P}_1)$

Proposition 1. Expression for $\Delta_\mathrm{f} \geq 0$:

$$\Delta_\mathrm{f} = \sum_{\hat{x}=1}^{M} \hat{P}_0(\hat{x})\, D\Big(P_0(x|\hat{x}) \Big|\Big| P_1(x|\hat{x})\Big)$$

$P_\theta(X|\hat{X}) = \frac{P_\theta(X)}{\hat{P}_\theta(\hat{X})} \mathbb{1}\{\hat{X} = \mathrm{f}(X)\}$ is the posterior of X given $\hat{X} = \mathrm{f}(X)$.

- Note that a good task-aware compression strategy combines X that have similar posteriors $P_\theta(X|\hat{X})$.

Optimal compressor:

- $\mathrm{f}^\star = \arg\max_\mathrm{f} D(\hat{P}_0||\hat{P}_1) = \arg\min_\mathrm{f} \Delta_\mathrm{f}$ s.t. $|\mathrm{f}| \leq M$;
- optimization over each possible f, which induces a partition of M sets over $\mathcal{X}$ (NP-hard).

Proposed Compressor Scheme

Optimal one-step compression from $\lvert\mathcal{X}\rvert$ to $\lvert\mathcal{X}\rvert - 1$:

- f combines $\{a, b\} \subset \mathcal{X}$ and the others $x \in \mathcal{X} \setminus \{a, b\}$ are one-to-one;
- i.e., $\mathrm{f}(a) = \mathrm{f}(b) = m \in \mathcal{M}$, $\mathrm{f}(i) = i \in \mathcal{M} \setminus \{m\}$;

Then,

$$\mathrm{f}^\star = \underset{\{a,b\} \subset \mathcal{X} : \mathrm{f}(a) = \mathrm{f}(b) = m}{\arg\min} \left\{ \hat{P}_0(m) D\Big(P_0(x|m) \Big|\Big| P_1(x|m)\Big) \right\}. \tag{1}$$

Our "KL-greedy" compressor:

- iteratively reduce the alphabet size by 1 at each step, until the compressed alphabet has size M;
- at each step, combine $\{a, b\}$ which minimize (1);
- note that this compressor can be determined in polynomial type.

Results

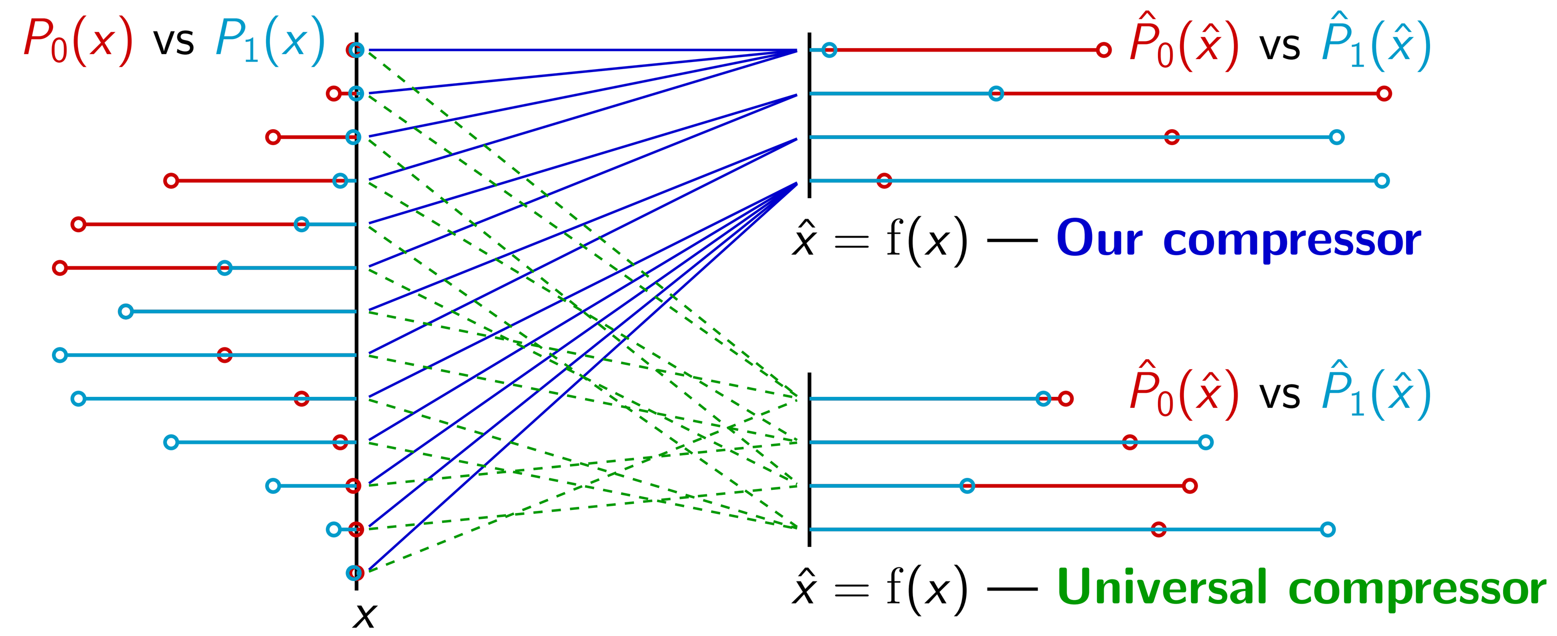
P_θ are shifted binomial distributions with different parameters.

Compare compression penalty Δ_f and empirical type-II error rate for:

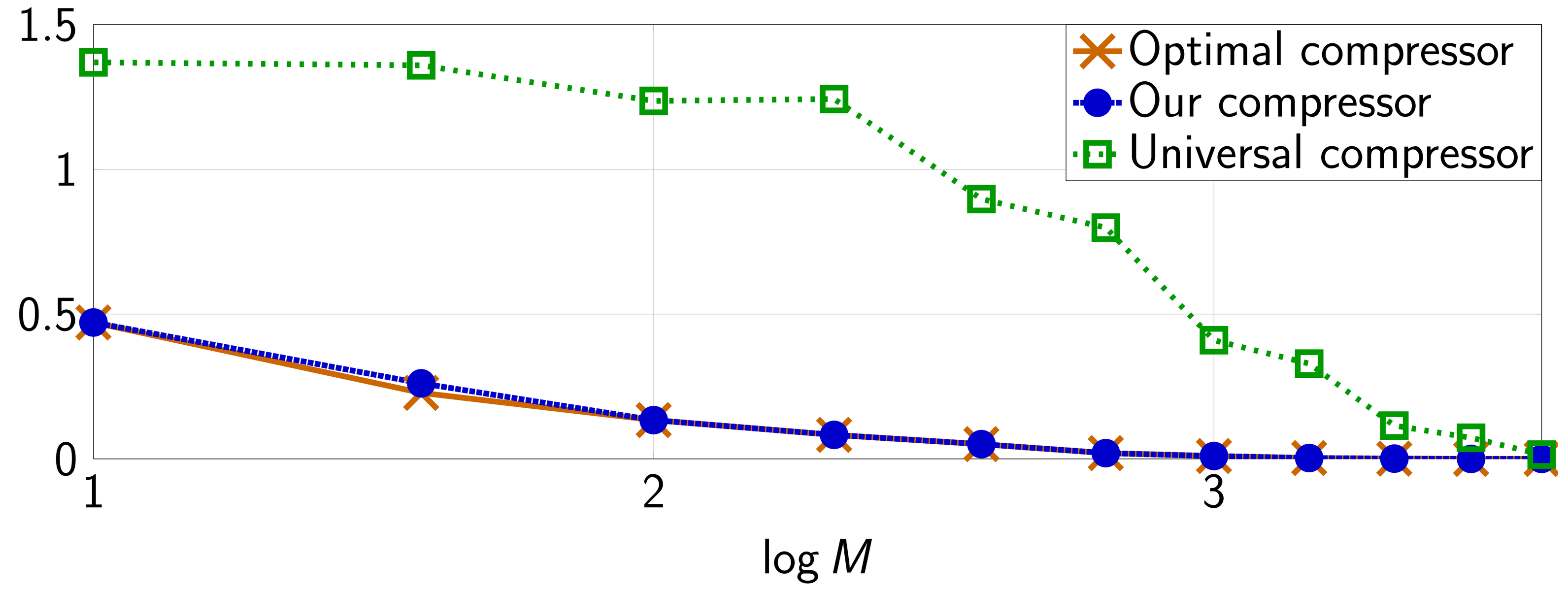
- optimal compressor $\mathrm{f}^\star$ — when feasible to compute, i.e, small $\lvert\mathcal{X}\rvert$;
- our KL-greedy compressor;
- universal compressor from [2], which is designed for reconstruction under log-loss distortion.

For the empirical type-II error rate, consider equal priors and $T = 1$.

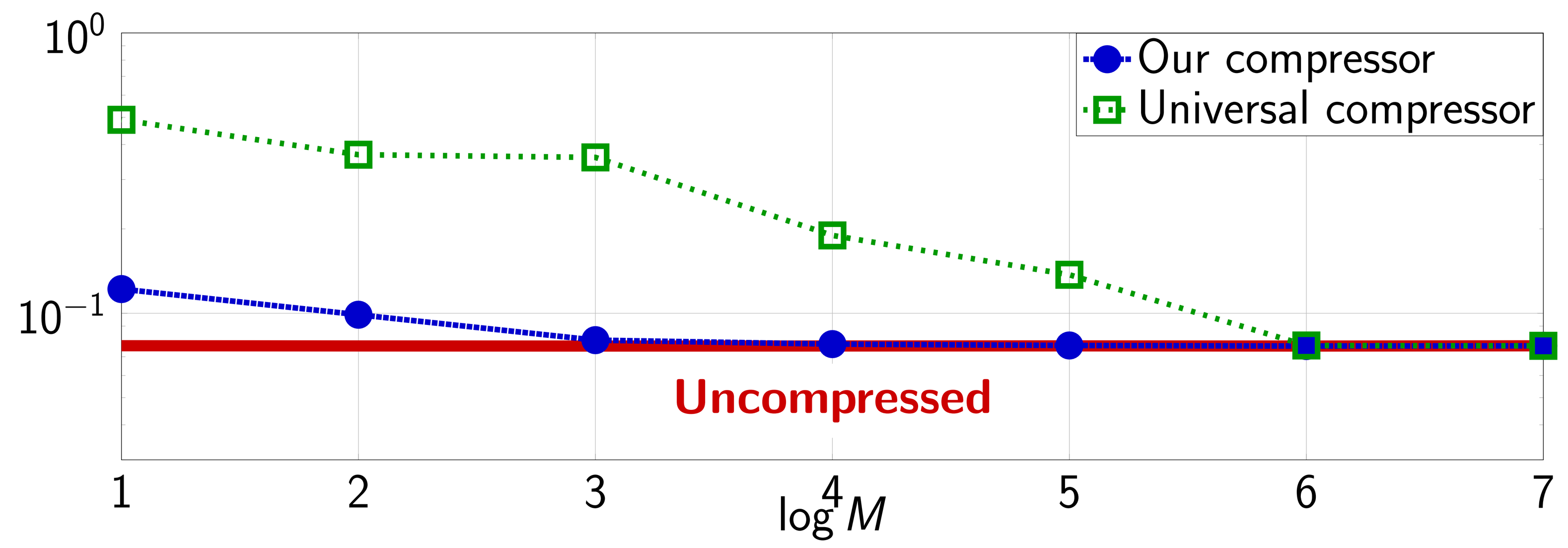
Results: Distributions and Compressor for $\lvert\mathcal{X}\rvert = 13$, $M = 4$



Results: Compression penalty $\Delta_\mathrm{f}(R)$ for $\lvert\mathcal{X}\rvert = 13$



Results: Type-II Error Rate for $\lvert\mathcal{X}\rvert = 256$, $n = 5$



Conclusions

- Formulation for the optimal compressor for hypothesis testing.
- Proposed the empirical "KL-greedy" compressor: it can be computed in polynomial time and preserves the *useful* information.
- Task-aware compression achieves error rate comparable to the uncompressed case for low rates.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[2] Y. Shkel, M. Raginsky, and S. Verdú, "Universal lossy compression under logarithmic loss," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 1157–1161.

[3] F. Carpi, S. Garg, and E. Erkip, "Single-shot compression for hypothesis testing," in *under review*. [Online]. Available: https://fabriziocarpi.github.io/