Multiterminal Source Coding under Logarithmic Loss

Thomas Courtade* and Tsachy Weissman†

Department of Electrical Engineering, *UCLA and †Stanford University

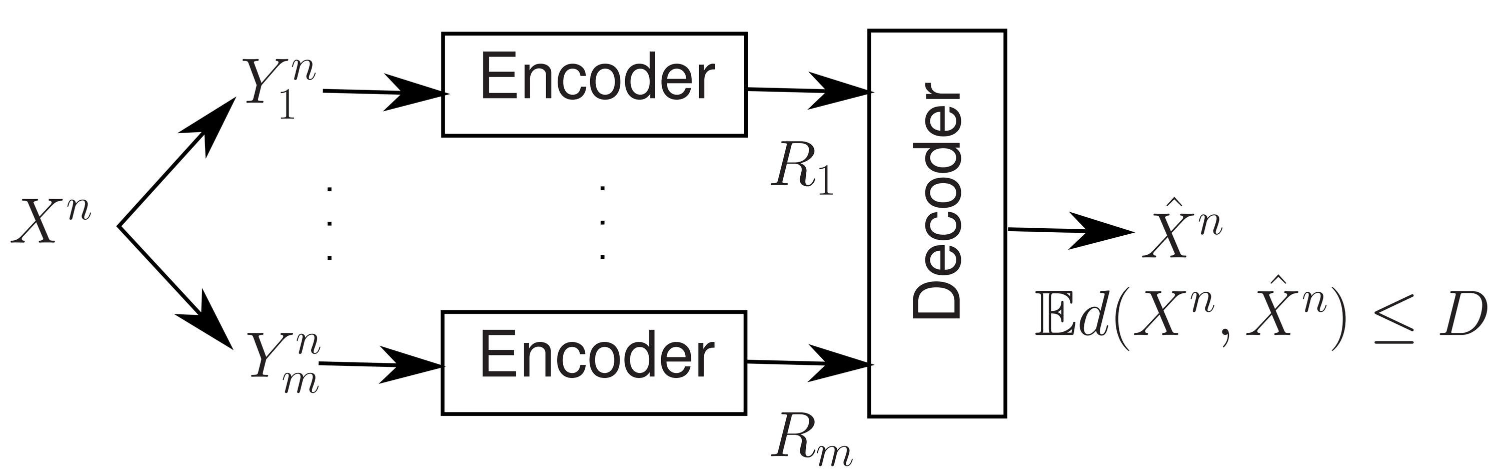
Introduction & Motivation

Unlike many classical settings (e.g., a binary source reproduced subject to a Hamming distortion constraint), several modern applications of rate distortion theory do not require deterministic decisions to be made at the decoder. For instance, an online recommendation system might guess a particular customer is male with probability 90% and female with probability 10%. Given this assignment of likelihoods, products can be recommended in a Bayes-optimal fashion to maximize profit.

Motivated by this, for a finite alphabet $\mathcal{X}$, define $\hat{\mathcal{X}}$ to be the set of probability measures on $\mathcal{X}$. Given $x \in \mathcal{X}$ and $\hat{x} \in \hat{\mathcal{X}}$, define the **logarithmic loss**:

$$d(x, \hat{x}) = \log\left(\frac{1}{\hat{x}(x)}\right).$$

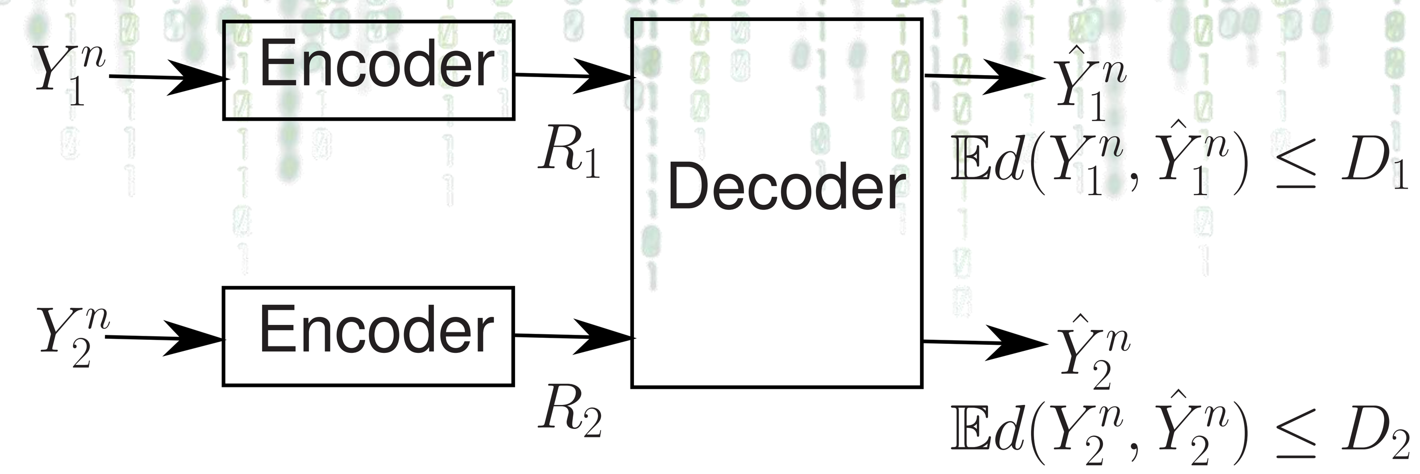
Two (Mostly) Open Problems



The CEO Problem

- In the CEO problem, the sources $Y_1, \ldots, Y_m$ are assumed to be conditionally independent given an unknown variable X. Prior to this work, only the Gaussian setting was completely solved [1, 2].

- The (two-encoder) multiterminal source coding problem extends Shannon's classical setup to two separate encoders which observe correlated sources. It is one of the simplest source coding networks which we don't entirely understand.



The Multiterminal Source Coding Problem

Previously, a complete characterization of the rate distortion region was only known in the Quadratic Gaussian setting [3].

Main Results

Theorem 1: $(R_1, \ldots, R_m, D)$ is achievable for the CEO problem under logarithmic loss if and only if

$$\sum_{i \in \mathcal{I}} R_i \geq I(Y_{\mathcal{I}}; U_{\mathcal{I}} | U_{\mathcal{I}^c}, Q) \text{ for all } \mathcal{I} \subset \{1, \ldots, m\}$$
$$D \geq H(X | U_1, \ldots, U_m, Q)$$

for some joint distribution of the form

$$p(x)p(q)\prod_{i=1}^{m} p(y_i|x)p(u_i|y_i, q).$$

Theorem 2: (R_1, R_2, D_1, D_2) is achievable for the multiterminal source coding problem under logarithmic loss if and only if

$$\begin{aligned}
R_1 &\geq I(Y_1; U_1 | U_2, Q)\\
R_2 &\geq I(Y_2; U_2 | U_1, Q)\\
R_1 + R_2 &\geq I(Y_1, Y_2; U_1, U_2 | Q)\\
D_1 &\geq H(Y_1 | U_1, U_2, Q)\\
D_2 &\geq H(Y_2 | U_1, U_2, Q)
\end{aligned}$$

for some joint distribution of the form

$$p(y_1, y_2)p(q)p(u_1|y_1, q)p(u_2|y_2, q).$$

Remark: The Berger-Tung inner bound is tight for both problems under Logarithmic Loss.

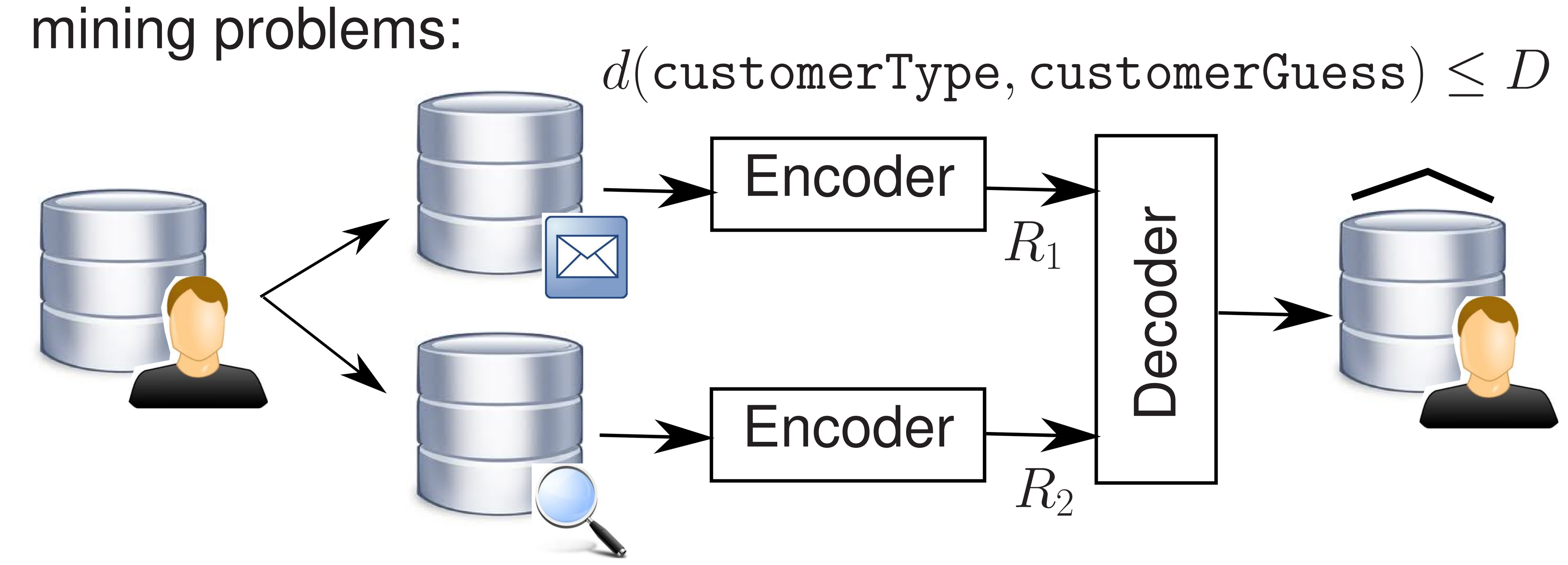
Proof Sketch

The proof of Theorem 2 consists of three main steps:

1. **Solve the CEO Problem**. Key Ideas: A Fano-type inequality for expected distortion, and a submodularity argument to show inner & outer bounds coincide.
2. **Couple the MTSC problem to a CEO Problem**. Key Idea: Define a new random variable X as follows
$$X = \begin{cases} (Y_1, 1) \text{ with probability } t \\ (Y_2, 2) \text{ with probability } 1 - t. \end{cases}$$
$Y_1 \leftrightarrow X \leftrightarrow Y_2$, so we can apply converse of Thm. 1.
3. **Tune the parameter** t. Key Idea: Each choice of $t \in [0, 1]$ in Step 2 gives an outer bound. Can prove the existence of a t^* for which the outer bound meets the Berger-Tung inner bound.

Applications

- The CEO problem is a good representation of data mining problems:



- Theorem 2 can be applied to list decoding, horse racing, and can also be used to obtain outer bounds for the general multiterminal source coding problem.

References

[1] V. Prabhakaran, D. Tse, and K. Ramachandran. "Rate region of the quadratic Gaussian CEO problem." Proc. ISIT, July 2004.

[2] Y. Oohama. "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder." Trans. IT, July 2005.

[3] A.B. Wagner, S. Tavildar, and P. Viswanath. "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem." Trans. IT, May 2008.

[4] T. A. Courtade and T. Weissman. "Multiterminal Source Coding under Logarithmic Loss." *CoRR*, abs/1110.3069, 2011.

Center for Science of Information

A National Science Foundation Science and Technology Center

soihub.org