Lecture 10, Part 1:

Channel Coding and Source/Channel Separation

Robert Bamler · Summer Term of 2023

These slides are part of the course *"Data Compression With and Without Deep Probabilistic Models"* taught at University of Tübingen. More course materials—including video recordings, lecture notes, and problem sets with solutions—are publicly available at `https://robamler.github.io/teaching/compress23/`.

Recall: Source Coding Theorem

- optimal expected bit rate of *lossless* compression: entropy $H[\mathbf{X}]$
- we proved that $H[\mathbf{X}]$ is both:
  - a lower bound: $\mathbb{E}[\text{bitrate}(\mathbf{X})] \geq H[\mathbf{X}] = \mathbb{E}[-\log P(\mathbf{X})] \quad \forall\,\text{lossless codes}$
  - achievable with negligible overhead: $\exists\,\text{lossless code} : \text{bitrate}(\mathbf{x}) < -\log P(\mathbf{x}) + 1 \;\forall\,\mathbf{x}$
- **Today:** *lossy* compression can have bit rates $< H[\mathbf{X}]$.

  But there is still a lower bound.

  1. lower bound: easy to prove
  2. achievability of lower bound: via a detour through channel coding

Lower Bound on the Bit Rate of Lossy Compression

- Encoder/decoder form a *Markov chain:*
- Problem 10.3: data processing inequality:

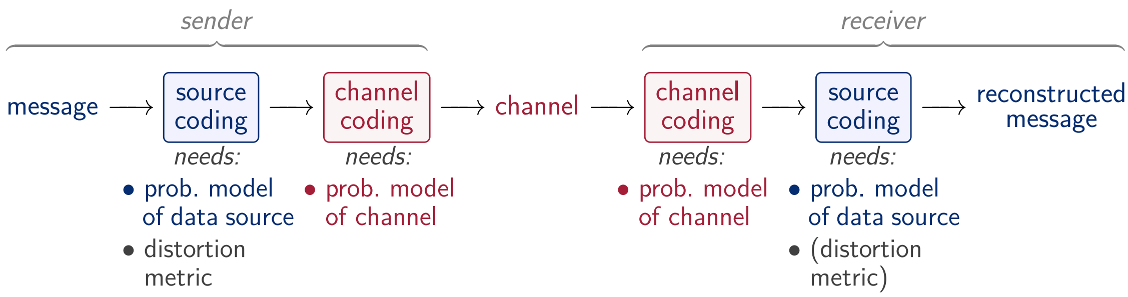
  $\forall$ Markov chains $X_1 \to X_2 \to X_3$:
- Thus, lower bound on expected bit rate:
  - consider data source $P(\mathbf{X})$ and fixed mapping $P(\mathbf{X}' \,|\, \mathbf{X})$ from messages to reconstructions;
  - encoder $P(\mathbf{S} \,|\, \mathbf{X})$ and decoder $P(\mathbf{X}' \,|\, \mathbf{S})$ satisfy:

Poll

- Assume we want to transmit a message $\mathbf{x} \in \{0,1\}^k$ of k independent and uniformly distributed bits over a channel.
- The channel can only transmit a string of $n < k$ bits without error.
- How many of the n bits in the message should we expect to be corrupted, in expectation?

Channel Coding

- **Recap from very first lecture:**



Error Correction: Intuition

- Assume we want to transmit a message $\mathbf{x} \in \{0,1\}^k$ of k independent and uniformly distributed bits over a channel.
- We have a channel that transmits bits $\in \{0,1\}$, but it flips each bit with some probability f (independently of each ohter).
- **Question 1:** what is the probability that the message is transmitted without error? For example, consider a 1 megabit message and error probability $f = 0.001\,\%$.
- **Question 2:** now consider a simple channel coding scheme: the each bit is transmitted three times; the receiver then takes a majority vote of the three bits. What is now the probability to receive the message without error in the above example?

(Noisy) Channel Coding Theorem

Claim: we can do a lot better than replicating each bit three times:

$$\mathbf{S} \xrightarrow[P(\mathbf{Y}|\mathbf{X})]{\text{channel encoder}} \mathbf{Y} \xrightarrow[\prod_{i=1}^n P(Y_i|X_i)]{\text{memoryless channel}} \mathbf{Y}' \xrightarrow[P(\mathbf{X}'|\mathbf{Y})]{\text{channel decoder}} \mathbf{S}'$$

- For a memoryless Channel $P(\mathbf{Y} \,|\, \mathbf{X}) = \prod_{i=1}^{n} P(Y_i \,|\, X_i)$, let the *channel capacity* be:

$$C := \sup_{P(X_i)} I_P(X_i; Y_i).$$

- Then: in the limit of long messages ($n \gg 1$), there exists a channel coding scheme that satisfies both of the following:
  - the ratio $\frac{k}{n}$ can be made arbitrarily close to C; and
  - the error probability $P(\mathbf{S} \neq \mathbf{S}')$ can be made arbitrarily small for all $\mathbf{s} \in \{0,1\}^k$.

Intuition: Block Error Correction

Prerequisite 1 of 2: Chebychev's Inequality

- Let X be a nonnegative (discrete or continuous) scalar random variable with a finite expectation $\mathbb{E}_P[X]$. Then:

$$P(X \geq \beta) \leq \frac{\mathbb{E}_P[X]}{\beta} \qquad \forall \beta > 0.$$

- **Proof:**

Prerequisite 2 of 2: Weak Law of Large Numbers

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

- Let $X_1, \ldots, X_n$ be independent random variables, all with the same expectation value $\mu := \mathbb{E}_P[X_i]$, and with the same (finite) variance $\sigma^2 := \mathbb{E}_P\big[(X_i - \mu)^2\big] < \infty$.
- Denote the *empirical mean* of all X_i as $\langle X_i \rangle_i := \frac{1}{n} \sum_{i=1}^n X_i$ (thus, $\langle X_i \rangle_i$ is itself a random variable).
- Then: $P\Big(\big|\langle X_i \rangle_i - \mu\big| \geq \beta\Big) \leq \frac{\sigma^2}{n\beta^2} \qquad \forall \beta > 0.$
- **Proof:**

Implications on Information Content

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

What are "Typical" Messages?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Last Slide: $P\left(\left|\frac{-\log_2 P(\mathbf{X})}{n} - H_P[X_i]\right| \geq \beta\right) \leq O\left(\frac{1}{n\beta^2}\right) \qquad \forall \beta > 0$

- Thus, for "most" long random messages, the information content per symbol is close to the entropy of a symbol.
- Define the *typical set* $T_{P(X_i),n,\beta}$ as the set of messages of length n whose information content per symbol deviates from the entropy of a symbol by less than some given threshold β:

Examples of Typical Sets

Size of the Typical Set

Application to Channel Coding

Understanding Joint Typicality

Random Channel Codes

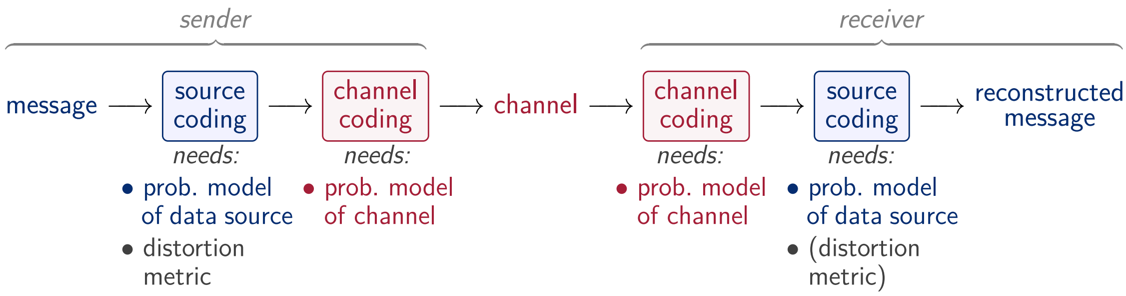
Proof

Proof of the Noisy Channel Coding Theorem

Recall: Poll

- Assume we want to transmit a message $\mathbf{x} \in \{0,1\}^k$ of k independent and uniformly distributed bits over a channel.
- The channel can only transmit a string of $n < k$ bits without error.
- How many of the n bits in the message should we expect to be corrupted, in expectation?

Application to Lossy Data Compression



Achievability of the Lower Bound (1 of 2)

Achievability of the Lower Bound (2 of 2)

Source/Channel Separation Theorem