Fall 2025

6.7480 - Information Theory: From Coding to Learning

Problem Set 9

Due: Dec 10, 2025, 10:00pm

`Prof. Yury Polyanskiy`

RULES:

- **Submit any 6 exercises. (If you submit more than 6, we will attempt to heuristically guess the best-6 to grade. If you want us avoid guessing, indicate which 6 you want to have graded.)**
- Each exercise is **15 points**.
- Note that the higher number of problems and 15 points per problem are due to the fact that this is a PSet that replaces the final exam.
- Submit your solution as **one file** (e.g., typed or scanned PDF) on Gradescope.
- **Collaboration** is allowed, but you have to state clearly who you collaborated with.
- AI written solutions are prohibited.

---

1 (Product source) Consider two independent stationary memoryless sources $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ with reproduction alphabets $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$, distortion measures $d_1 : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_+$ and $d_2 : \mathcal{Y} \times \hat{\mathcal{Y}} \to \mathbb{R}_+$, and rate-distortion functions R_X and R_Y, respectively. Now consider the stationary memoryless product source $Z = (X, Y)$ with reproduction alphabet $\hat{\mathcal{X}} \times \hat{\mathcal{Y}}$ and distortion measure $d(z, \hat{z}) = d_1(x, \hat{x}) + d_2(y, \hat{y})$.

(a) Show that

$$I(X, Y; \hat{X}, \hat{Y}) \geq I(X; \hat{X}) + I(Y; \hat{Y})$$

provided that X and Y are independent.

(b) Show that the rate-distortion function of Z is related to that of X and Y via the *inf-convolution*, i.e.

$$R(D) = \inf_{0 \leq D_1 \leq D} R_X(D_1) + R_Y(D - D_1).$$

(c) How do you build an optimal lossy compressor for Z using optimal lossy compressors for X and Y?

2 (Noisy source-coding; also remote source-coding) Consider the problem of compressing i.i.d. sequence X^n under separable distortion metric d. Now, however, compressor does not have direct access to X^n but only to its noisy version Y^n obtained over a stationary memoryless channel $P_{Y|X}$ (i.e. $(X_i, Y_i) \stackrel{iid}{\sim} P_{X,Y}$ for a fixed $P_{X,Y}$ and encoder is a map $f : \mathcal{Y}^n \to [M]$). Show that the rate-distortion function is

$$R(D) = \min \left\{ I(Y; \hat{X}) : \mathbb{E}\left[d(X, \hat{X})\right] \leq D, X \to Y \to \hat{X} \right\},$$

where minimization is over all $P_{\hat{X}|Y}$. (Hint: define $\tilde{d}(y, \hat{x}) \triangleq \mathbb{E}\left[d(X, \hat{x}) | Y = y\right]$.)

3 (Non-asymptotic rate-distortion) Our goal is to show that the convergence to $R(D)$ happens much faster than that to capacity in channel coding. Consider binary uniform $X \sim \text{Ber}(1/2)$ with Hamming distortion.

(a) Show that there exists a lossy code $X^n \to W \to \hat{X}^n$ with M codewords and

$$\mathbb{P}[d(X^n, \hat{X}^n) > D] \leq (1 - p(nD))^M \ ,$$

where

$$p(s) = 2^{-n} \sum_{j=0}^{s} \binom{n}{j} .$$

(b) Show that there exists a lossy code with M codewords and

$$\mathbb{E}\left[d(X^n, \hat{X}^n)\right] \leq \frac{1}{n} \sum_{s=0}^{n-1} (1 - p(s))^M . \tag{1}$$

(c) Show that there exists a lossy code with M codewords and

$$\mathbb{E}\left[d(X^n, \hat{X}^n)\right] \leq \frac{1}{n} \sum_{s=0}^{n-1} e^{-Mp(s)} . \tag{2}$$

(Note: For $M \approx 2^{nR}$, numerical evaluation of (1) for large n is challenging. At the same time (2) is only slightly looser.)

(d) For $n = 10, 50, 100$ and 200 compute the upper bound on $\log M^*(n, 0.11)$ via (2). Compare with the lower bound

$$\log M^*(n, D) \geq nR(D) . \tag{3}$$

4 Let $X \sim \text{Unif}[\pm 1]$ and reconstruction $\hat{X} \in \mathbb{R}$ with $d(x, \hat{x}) = (x - \hat{x})^2$. Show that rate-distortion function is given in parametric form by

$$R = \log 2 - h(p), \qquad D = 4p(1-p), \qquad p \in [0, 1/2]$$

and that for any distortion level optimal vector quantizer is only taking values $\pm(1 - 2p)$ (Hint: you may find Exercise I.64(b) useful). Compare with the case of $\hat{X} \in \{\pm 1\}$, for which we have shown $R(D) = \log 2 - h(D/4)$, $D \in [0, 2]$.

5 (Water-filling solution)

(a) Let $0 \prec \Delta \preceq \Sigma$ be positive definite matrices. For $S \sim \mathcal{N}(0, \Sigma)$, show that

$$\inf_{P_{\hat{S}|S} : \mathbb{E}[(S - \hat{S})(S - \hat{S})^\top] \preceq \Delta} I(S; \hat{S}) = \frac{1}{2} \log \frac{\det \Sigma}{\det \Delta} .$$

(Hint: for achievability, consider $S = \hat{S} + Z$ with $\hat{S} \sim \mathcal{N}(0, \Sigma - \Delta) \perp\!\!\!\perp Z \sim \mathcal{N}(0, \Delta)$ and apply Example 3.5; for converse, follow the proof of Theorem 26.2.)

(b) Prove the following extension of (26.3): Let $\sigma_1^2, \ldots, \sigma_d^2$ be the eigenvalues of Σ. Then

$$\inf_{P_{\hat{S}|S}:\mathbb{E}[\|S-\hat{S}\|_2^2]\le D} I(S;\hat{S}) = \frac{1}{2}\sum_{i=1}^{d} \log^+ \frac{\sigma_i^2}{\lambda}$$

where $\lambda > 0$ is such that $\sum_{i=1}^d \min\{\sigma_i^2, \lambda\} = D$. This is the counterpart of the solution in Theorem 20.14.

(Hint: First, using the orthogonal invariance of distortion metric we can assume that Σ is diagonal. Next, apply the same single-letterization argument for (26.3) and solve $\min_{\sum D_i = D} \frac{1}{2}\sum_{i=1}^d \log^+ \frac{\sigma_i^2}{D_i}$.)

6 (Distribution estimation in TV) Given $X_1, \ldots, X_n$ drawn independently from a distribution P over $[k]$, we show that the minimax rate for estimating P with respect to the total variation loss is

$$R^*_{\mathrm{TV}}(k,n) \triangleq \inf_{\hat{P}} \sup_{P\in\mathcal{P}_k)} \mathbb{E}_P[\mathrm{TV}(\hat{P},P)] \asymp \sqrt{\frac{k}{n}} \wedge 1, \qquad \forall k\ge 2, n\ge 1, \tag{4}$$

(a) Show that the MLE coincides with the empirical distribution.

(b) Show that the MLE achieves the RHS of (4) within constant factors. (Hint: either apply (7.58) plus Pinsker's inequality, or directly use the variance of empirical frequencies.)

(c) Establish the minimax lower bound. (Hint: apply Assouad's lemma, or Fano's inequality (with volume method or explicit construction of packing), or the mutual information method directly.)

7 (Covering radius in Hamming space) In this exercise we prove (27.9), namely, for any fixed $0 \le D \le \frac{1}{2}$, as $n \to \infty$,

$$N(\mathbb{F}_2^n, d_H, Dn) = 2^{n(1-h(D))+o(n)},$$

where $h(\cdot)$ is the binary entropy function.

(a) Prove the lower bound by invoking the volume bound in Theorem 27.3 and the large-deviations estimate in Example 15.1.

(b) Prove the upper bound using probabilistic construction and a similar argument to (25.8).

(c) Show that for $D \ge \frac{1}{2}$, $N(\mathbb{F}_2^n, d_H, Dn) \le 2$ – cf. Ex. V.15(a).

8 (Covering ℓ_p-ball with ℓ_q-balls)

(a) For $1 \le p < q \le \infty$, prove the following bound on the metric entropy of the unit ℓ_p-ball with respect to the ℓ_q-norm:

$$\log M(B_p, \|\cdot\|_q, \epsilon) \asymp_{p,q} \begin{cases} d\log\left(\frac{e}{\epsilon^s d}\right) & \epsilon \le d^{-1/s} \\ \frac{1}{\epsilon^s}\log(e\epsilon^s d) & \epsilon \ge d^{-1/s} \end{cases}, \qquad \frac{1}{s} \triangleq \frac{1}{p} - \frac{1}{q}.$$

(Hint: for small ϵ, apply the volume calculation in (27.15)–(27.16) and the formula in (27.13); for large ϵ, proceed as in the proof of Theorem 27.7 by applying the quantization argument and the Gilbert-Varshamov bound of Hamming spheres.)

(b) What happens when $p > q$?

9 (Covering and discrepancy theory) Let $r = n/2 - 3\sqrt{n}$ and consider m Hamming balls $B(c_1, r), \ldots, B(c_m, r)$ covering the Hamming space $\{0,1\}^n = \cup_j B(c_j, r)$. We will show that $m > n$.

(a) Show that rate-distortion lower bound only gives $m \gtrsim 1$.

(b) Show that existence of covering implies existence of a matrix $A \in \{\pm 1\}^{m \times n}$ such that

$$\min_{\epsilon \in \{\pm 1\}^n} \|A\epsilon\|_\infty \geq 6\sqrt{n}\,.$$

(c) Conclude the proof by appealing to *Spencer's six-sigma theorem:* For any $A \in [-1,1]^{n \times n}$ there exists $\epsilon \in \{\pm 1\}^n$ such that $\|A\epsilon\|_\infty < 6\sqrt{n}$.

Note: Hadamard code covers the space by n balls with radius $r' = n/2 - \sqrt{n/2}$. The reverse implication also holds: a covering radius lower bound $r > n/2 - \Omega(\sqrt{n})$ for $m \asymp n$ implies Spencer's theorem.