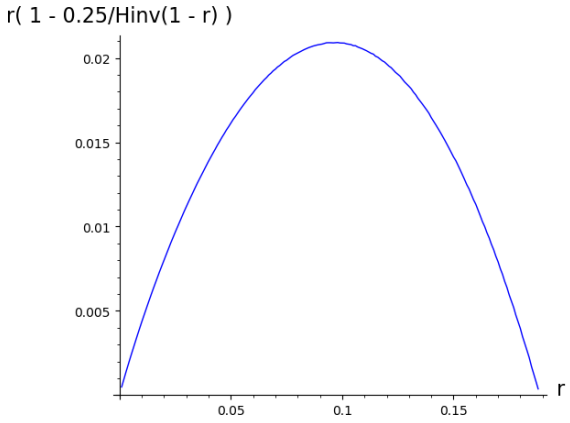
Class 7 Exercises

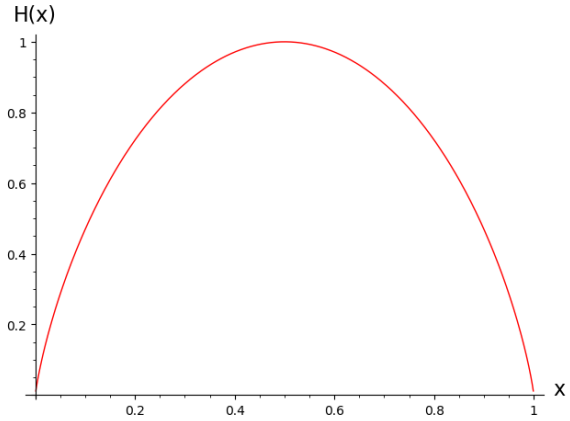
CS250/EE387, Winter 2025

1. What is (approximately) the rate of the best explicit binary code that we know exists with distance 1/4? Compare this to (approximately) the rate of the best non-explicit binary code that we know exists with distance 1/4.

   If it helps, here is a plot of the function $\phi(r) = r \cdot \left(1 - \frac{1/4}{H^{-1}(1-r)}\right)$:



   And here is a plot of the binary entropy function:



2. With the Zyablov bound, we saw (finally) a construction of an "explicit" asymptotically good binary code using concatenated codes. However, the definition of "explicit" may not have been super-satisfying. In this exercise we'll see a construction with a more satisfying definition of "explicit."

   (a) Explain why the construction from the videos might not be super satisfying in terms of explicitness. (Recall that "explicit" meant that "there exists a polynomial-time algorithm to find it." What could we hope for instead?)

   (b) Fix a parameter k. Let $Q = 2^k$. We can identify the finite field $\mathbb{F}_Q$ with the vector space $\mathbb{F}_2^k$. These aren't the same thing, but it turns out that they have the same additive structure. That is, there is some map $\varphi : \mathbb{F}_2^k \to \mathbb{F}_Q$ so that

   $$\varphi(\mathbf{0}) = 0 \qquad \text{and} \qquad \varphi(\mathbf{x}+\mathbf{y}) = \varphi(\mathbf{x}) + \varphi(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{F}_2^k.$$

   For $\alpha \in \mathbb{F}_Q$, with $\alpha \neq 0$, consider the linear code $\mathcal{C}_\alpha \subset \mathbb{F}_2^{2k}$ of dimension k and length $2k$ over $\mathbb{F}_2$ given by the encoding map

   $$\mathbf{x} \mapsto \mathbf{x} \circ \varphi^{-1}(\varphi(\mathbf{x}) \cdot \alpha),$$

   where $\circ$ denotes concatenation.

   There is no question for this part, just make sure you understand the construction. Notice that there are $2^k - 1$ different codes C_α that you can create like this.

   (c) Let $\varepsilon > 0$, and let $\eta = 2^{-2\varepsilon k}$. Show that at least a $1 - \eta$ fraction of the α's in $\mathbb{F}_2$ have

   $$\text{dist}(\mathcal{C}_\alpha) \geq H_2^{-1}(1/2 - \varepsilon).$$

   Hint. Consider a fixed nonzero $\mathbf{y} \in \mathbb{F}_2^{2k}$ with weight less than $2k \cdot H_2^{-1}(1/2 - \varepsilon)$. How many such vectors $\mathbf{y}$ are there? How many different codes $\mathcal{C}_\alpha$ can each vector $\mathbf{y}$ lie in?

   (d) Let $Q = 2^k$ as before. Consider the code $\mathcal{C}$ (called the *Justesen code*) formed as follows:

   - Let $\mathcal{RS}$ be a Reed-Solomon code over $\mathbb{F}_Q$ with length $N = Q - 1$, evaluation points $\mathbb{F}_Q \setminus \{0\}$, and rate R. Thus, each symbol of $\mathcal{RS}$ is associated with an element $\alpha \in \mathbb{F}_Q \setminus \{0\}$.
   - For each $\alpha \in \mathbb{F}_Q \setminus \{0\}$, let $\mathcal{C}_\alpha$ be as in the previous part.
   - Consider the "concenated-like" code where the inner code is *different* for each symbol: we concatenate the α'th symbol of a codeword in $\mathcal{RS}$ with the inner code $\mathcal{C}_\alpha$.

   Show that the rate of this code is at least $R/2$, and that the distance is at least $(1 - R - \eta) \cdot H_2^{-1}(1/2 - \varepsilon)$.

   (e) Does the Justesen code meet your definition of "satisfyingly explicit" from part (a)? Why or why not?

   Note: If you haven't seen the correspondence between $\mathbb{F}_{2^k}$ and $\mathbb{F}_2^k$ (as per the map φ above) before, it may be tricky to answer this question precisely. In that case, just try to think about whether it seems plausible and then move on.

   (f) Does this construction give you an asymptotically good code of distance, say, 1/4 (c.f. Exercise 1)? How about for distance 1/20? How does this compare to the Zyablov/GV bound? (Better or worse?) (**Note:** for distance 1/20, the Zyablov bound is about 0.27. The binary entropy of 1/20 is about 0.29).

   Hint. Since the question is about asymptotics, you can set $\varepsilon, \eta = 0$ in your answer to (d). (The reason is that we can choose $\varepsilon > 0$ to be arbitrarily small; and then suppose that k is big enough compared to $1/\varepsilon$ that $\eta = 2^{-2k\varepsilon}$ is also arbitrarily small.)

   (g) **(Bonus).** How would you modify the construction above to achieve a better trade-off? Can you match the Zyablov bound this way?

   (h) **(Bonus 2).** Does the algorithm that we saw for decoding concatenated codes work for our Justesen code?

Stuff below here is super-bonus, we definitely won't get to it in class. The lecture notes promised that "in class" we'd see a construction of a code near the GV bound in time $2^{O(n)}$ (rather than $2^{O(n^2)}$)...but then we decided that the above would be more fun. If you want to know how to construct such a code, work through the steps below!

3. (**Deterministic codes on the GV bound in time $2^{O(n)}$.**) In the lecture videos/notes, we said that it is open (in most parameter regimes) to find an efficient deterministic construction of an asymptotically good code near the GV bound, even though "most" linear codes lie close to this bound.

   (0) Fix some constant δ and let $\varepsilon > 0$. Describe a straightforward algorithm that finds an asymptotically good code of length n near the GV bound (that is, with dimension at least $(1 - H(\delta) - \varepsilon)n$ and distance at least δ) in time $2^{O(n^2)}$. In the big-Oh notation, we treat ε, δ as constants and n as growing.

   However, we can do a little bit better than $2^{O(n^2)}$ (and this may be useful soon for using concatenated codes to get an explicit construction of an asymptotically good code...).

   In this exercise we'll prove the following theorem:

   Theorem 1. *There is a deterministic algorithm that finds a binary linear code $\mathcal{C} \subset \mathbb{F}_2^n$ with rate $R = 1 - H_2(\delta) - o(1)$ in time $2^{O(n)}$.*

   Fix constants $\varepsilon, \delta > 0$. Suppose that $k \geq 1 - H_2(\delta) - \varepsilon$. In the following, we will come up with a distribution $\mathcal{D}$ on matrices $G \in \mathbb{F}_2^{n \times k}$ so that

   (i) A matrix $G \sim \mathcal{D}$ can be sampled using $O(n)$ bits of randomness.

   (ii) A matrix $G \sim \mathcal{D}$ is full rank with probability at least $2/3$.

   (iii) Let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be a (random) code with generator matrix $G \sim \mathcal{D}$; then $\mathcal{C}$ has distance at least δ with probability at least $2/3$.

   (a) Explain why coming up with such a distribution $\mathcal{D}$ would prove the theorem.

   (b) Define a distribution $\mathcal{D}$ on matrices G by letting G be a random *Toeplitz matrix*. That is, G is of the form

   $$G = \begin{pmatrix} X_0 & X_1 & \cdots & X_{k-1} \\ X_k & X_0 & X_1 & \ddots \\ X_{k+1} & X_k & X_0 & \ddots \\ \vdots & & & \\ X_{n+k-2} & & & \end{pmatrix}.$$

   where $X_0, \ldots, X_{n+k-2}$ are i.i.d. uniform random variables in $\mathbb{F}_2$.
   (There is no question here, just understand the distribution).

   (c) Show that, for any $x \in \mathbb{F}_2^k$, Gx is uniformly distributed when $G \sim \mathcal{D}$.

   (d) Show that (i), (ii), and (iii) are satisfied by $\mathcal{D}$.