

Topics in Error-Correcting Codes

MIT 6.S953 — Problem Set 2

Due: October 28 at 6 PM.

Presentation instructions: Solve at least one of Problems 1–2 and at least two of Problems 3–6. Only a presentation is required; no submission is required. [Schedule your presentation here](#). Problem 0 is for practice only.

Throughout, $\Delta(u, v)$ and $\text{wt}(u)$ denote Hamming distance and Hamming weight. A linear code with parameters $(n, k, d)_q$ has rate k/n and relative distance d/n ; a general code $\mathcal{C} \subseteq \Sigma^n$ has rate $\log_{|\Sigma|} |\mathcal{C}|/n$.

Problem 0: Spectral graph theory (practice)

Adapted from Salil Vadhan, [Pseudorandomness, Problem 2.9](#), pp. 29–30.

Recall the spectral theorem: Every real symmetric matrix has real eigenvalues and an orthonormal basis of eigenvectors.

Let G be an undirected d -regular multigraph on $n \geq 2$ vertices, with $d \geq 1$; loops are allowed. Write $M = A/d$ for its random-walk matrix, where A is symmetric with row sums d . Order its eigenvalues as $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, and put $\gamma(G) = 1 - \max_{i \geq 2} |\lambda_i|$. Prove:

- (a) The spectrum of M lies in $[-1, 1]$.
- (b) The multiplicity of eigenvalue 1 exceeds one exactly when G is disconnected.
- (c) For connected G , bipartiteness is equivalent to having eigenvalue -1 .
- (d) Connectivity implies $\lambda_2 \leq 1 - 1/\text{poly}(n, d)$. *Hint:* Establish the variational identity

$$1 - \lambda_2 = \min_{\substack{x \perp \mathbf{1} \\ \|x\|_2=1}} \frac{1}{2d} \sum_{i,j} A_{ij} (x_i - x_j)^2.$$

- (e) If G is connected and nonbipartite, then $\gamma(G) \geq 1/\text{poly}(n, d)$.
- (f) With D denoting the diameter, strengthen part (d) to $1 - \lambda_2 = \Omega(1/(dDn))$. Deduce $\gamma(G) = \Omega(1/(d^2n^2))$ for connected, nonbipartite G .

Problem 1: Interactive coding without chunk numbers

Recall the computationally efficient interactive coding scheme from Lecture 8. Each block header contains the chunk number, a hash of the transcript so far, and the hash key.

Consider a simplified variant in which the block headers omit the chunk number and contain only the transcript hash and its key. The parties advance to the next block when their hashes agree and go back one block when their hashes disagree.

Does this variant give a good interactive coding scheme, with the guarantees proved in class? Either prove that it does or give an adversarial attack. If you give an attack, specify the corruptions and explain why the resulting execution violates the desired guarantee.

Problem 2: Lower triangular and Toeplitz matrices in tree codes

Recall Schulman's tree-code construction from class. A random lower triangular Toeplitz matrix M is chosen, and the sequence of labels along a path $b = (b_1, \dots, b_n)^T$ is given by Mb , using the alphabets and arithmetic from class. Recall that a matrix is *Toeplitz* if its entries are constant along each diagonal.

- What goes wrong if M is not lower triangular? Explain which requirement of a tree code can fail.
- What goes wrong in the analysis from class if M is lower triangular but is not Toeplitz? Identify the step that uses the Toeplitz condition and explain why that argument no longer applies.

Problem 3: Generalized Wozencraft ensembles

Parts (a)–(c) are adapted from Madhu Sudan's CS 229r (Fall 2023), [Problem Set 2](#), [Problem 3](#).

Fix a basis identifying $\mathbb{F}_{2^t}$ with $\mathbb{F}_2^t$ as vector spaces. The usual Wozencraft ensemble consists of the binary linear codes

$$C_\alpha = \{(x, \alpha x) : x \in \mathbb{F}_{2^t}\}, \quad \alpha \in \mathbb{F}_{2^t},$$

each with block length $n = 2t$ and message length $k = t$. Let $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$, with $h^{-1} : [0, 1] \rightarrow [0, 1/2]$. You may use the Hamming-ball estimate

$$\sum_{j=0}^{\lfloor \delta n \rfloor} \binom{n}{j} \leq 2^{nh(\delta)} \quad (0 \leq \delta \leq 1/2).$$

- Rates $1/\ell$.** For integers $\ell > 1$ and $t \geq 1$, construct a family $\{C_i\}_{i \in I}$ of binary linear codes of length ℓt and message length t , with

$$|I| = 2^{(\ell-1)t}, \quad C_i \cap C_j = \{0^{\ell t}\} \quad \text{for } i \neq j.$$

For every fixed $0 < \delta < h^{-1}(1 - 1/\ell)$, prove that the fraction of codes with minimum distance at most $\delta \ell t$ is at most

$$2^{\ell t h(\delta) - (\ell-1)t}.$$

Thus, as $t \rightarrow \infty$ with ℓ fixed, almost every member has relative distance greater than δ .

- Rates $1 - 1/\ell$: a counting criterion.** For integers $\ell > 1$ and $t \geq 1$, suppose a family of binary linear codes $\{C_\alpha\}_{\alpha \in A}$ satisfies

$$C_\alpha \subseteq \mathbb{F}_2^{\ell t}, \quad |C_\alpha| = 2^{(\ell-1)t}, \quad |A| = (2^t)^{\ell-1},$$

and each nonzero vector c belongs to at most $(2^t)^{\ell-2}$ members. Show that, for every fixed $0 < \delta < h^{-1}(1/\ell)$, the fraction of members with minimum distance at most $\delta \ell t$ is at most

$$2^{\ell t h(\delta) - t}.$$

Conclude that this family also attains relative distances arbitrarily close to the Gilbert–Varshamov bound as t grows.

- Constructing the high-rate ensemble.** Construct a family $\{C_\alpha\}_{\alpha \in A}$ satisfying all the conditions in part (b). *Hint:* Over $\mathbb{F}_{2^t}$, describe a systematic linear code with $\ell - 1$ message symbols and one parity symbol. Count the choices of its coefficients compatible with a fixed nonzero word.

Problem 4: Expanders, samplers, and AEL distance amplification

The sampler and AEL portions are adapted from the practice problem in Madhu Sudan's CS 229r (Fall 2023), [Problem Set 5, pp. 1–2](#).

Parts (a) and (d) are optional. You may use their conclusions in the other parts without proving them.

Let $B = (L, R, E)$ be a d -regular bipartite multigraph with $|L| = |R| = n$. Count edges with multiplicity. For $S \subseteq L$ and $T \subseteq R$, let $e(S, T)$ count edges between them, and write $e(u, T) = e(\{u\}, T)$. Define the normalized biadjacency matrix $M \in \mathbb{R}^{L \times R}$ by $M_{uv} = e(u, \{v\})/d$. Assume

$$\|Mx\|_2 \leq \lambda \|x\|_2 \quad \text{whenever } x \perp \mathbf{1}, \quad 0 \leq \lambda < 1.$$

This is a singular-value condition on the biadjacency matrix.

- (a) **Optional: expander mixing lemma.** Prove that, for all $S \subseteq L$ and $T \subseteq R$,

$$\left| e(S, T) - \frac{d|S||T|}{n} \right| \leq \lambda d \sqrt{|S||T| \left(1 - \frac{|S|}{n}\right) \left(1 - \frac{|T|}{n}\right)}.$$

Hint: Subtract the constant components of the two indicator vectors and apply Cauchy–Schwarz.

- (b) **A basic expansion consequence.** For nonempty $S \subseteq L$, put $s = |S|/n$ and let $\Gamma(S) \subseteq R$ be its set of neighbors. Show that

$$\frac{|\Gamma(S)|}{n} \geq \frac{s}{s + \lambda^2(1 - s)}.$$

Hint: Apply part (a) to S and $R \setminus \Gamma(S)$.

- (c) **Spectral expansion gives sampling.** For $0 < \varepsilon < 1$, call B an ε -sampler if every $T \subseteq R$ satisfies

$$\frac{1}{n} \left| \left\{ u \in L : \left| \frac{e(u, T)}{d} - \frac{|T|}{n} \right| \geq \varepsilon \right\} \right| \leq \varepsilon.$$

For $\mu = |T|/n$, prove the variance bound

$$\frac{1}{n} \sum_{u \in L} \left(\frac{e(u, T)}{d} - \mu \right)^2 \leq \lambda^2 \mu(1 - \mu).$$

Deduce that $\lambda \leq \varepsilon^{3/2}$ suffices for B to be an ε -sampler.

Hint: Write the vector of neighborhood densities as $M\mathbf{1}_T$ and center $\mathbf{1}_T$ by subtracting $\mu\mathbf{1}$. Apply the spectral bound, followed by Markov's inequality for the squared deviation.

- (d) **Optional: random regular graphs are samplers.** Sample a (d, d) -regular bipartite graph with n vertices on each side as in Gallager's ensemble shown in class. Prove that for every fixed $\varepsilon > 0$, some $d = O(\varepsilon^{-3})$, independent of n , makes the graph an ε -sampler with probability tending to one as $n \rightarrow \infty$.

- (e) **Alon–Edmonds–Luby distance amplification.** Fix $0 < \varepsilon < 1/2$. Let C_{out} be a code of rate at least $1 - 2\varepsilon$ and relative distance at least ε over the alphabet $\mathbb{F}_q^{k_{\text{in}}}$. Let C_{in} be a code over $\mathbb{F}_q$ of block length $n_{\text{in}} = d$, rate R , and relative distance δ . Consider the composition

$$\mathbb{F}_q^{k_{\text{out}}} \xrightarrow{C_{\text{out}}} (\mathbb{F}_q^{k_{\text{in}}})^{n_{\text{out}}} \xrightarrow{C_{\text{in}}} (\mathbb{F}_q^d)^{n_{\text{out}}} \xrightarrow{\text{sampler } B} (\mathbb{F}_q^d)^{n_{\text{out}}}.$$

Here C_{in} is applied to each outer symbol, and B is a d -regular $(\varepsilon/2)$ -sampler with n_{out} vertices on each side. Place each inner codeword on the d edges incident to its left vertex, then collect the incoming symbols at each right vertex into one output block in a fixed order.

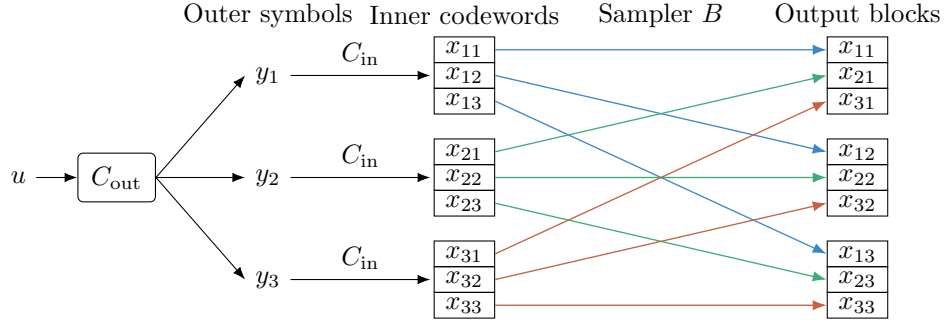


Illustration with three blocks and $d = 3$. Each wire carries one $\mathbb{F}_q$ -symbol; each output block is a symbol over $\mathbb{F}_q^d$.

Show that the resulting code over $\mathbb{F}_q^d$ has block length n_{out} , rate at least $(1 - 2\varepsilon)R$, and relative distance at least $\delta - \varepsilon/2$. Thus its rate and relative distance are at least $R - O(\varepsilon)$ and $\delta - O(\varepsilon)$, respectively.

Hint: For two distinct messages, let S be the left vertices where their outer encodings differ and let T be the right vertices where their final blocks differ. Every $u \in S$ has at least δd incident edges entering T . Explain where the sampler property is used and why regrouping the coordinates changes the relative distance.

Problem 5: BCH codes: distance and rate

Adapted from Guruswami–Rudra–Sudan, *Essential Coding Theory, Exercise 5.11*, pp. 103–104.

Let $m \geq 2$, $q = 2^m$, and $n = q - 1$. Let $\alpha = (\alpha_1, \dots, \alpha_n)$ enumerate the nonzero elements of $\mathbb{F}_q$. For an integer $0 \leq k \leq n$, define the binary BCH code

$$C_{\text{BCH}} = C_{\text{BCH}}(m, k) = \text{RS}_q[\alpha, k] \cap \mathbb{F}_2^n.$$

Here $\text{RS}_q[\alpha, k]$ evaluates polynomials of degree less than k at the points in α ; C_{BCH} consists of the codewords whose coordinates all lie in $\mathbb{F}_2$.

- (a) Let $d = n - k + 1$. Show that C_{BCH} is binary linear, has minimum distance at least d , and has dimension at least

$$n - (d - 1) \log_2(n + 1).$$

Hint: Use the parity-check characterization of Reed–Solomon codes from Exercise 5.9 in the text-book.

- (b) Strengthen the dimension lower bound to

$$n - \left\lceil \frac{d - 1}{2} \right\rceil \log_2(n + 1).$$

Hint: Some parity checks become redundant when the coordinates are restricted to $\mathbb{F}_2$.

- (c) Which code from class is C_{BCH} when $d = 3$?
- (d) For even d , show how to construct a length- n binary linear code of minimum distance at least d and dimension at least

$$n - \left(\frac{d}{2} - 1 \right) \log_2(n + 1) - 1$$

by imposing the global parity check $c_1 + \dots + c_n = 0$ on a suitable BCH code.

- (e) Deduce that whenever $n = 2^m - 1$ and $2 \leq d < n/\log_2(n+1)$, a binary linear code with parameters $(n, k', d')_2$ can be constructed with

$$d' \geq d, \quad k' \geq n - \left\lfloor \frac{d-1}{2} \right\rfloor \log_2(n+1) - 1.$$

- (f) Show that, for fixed d , the coefficient $\lfloor (d-1)/2 \rfloor$ of $\log_2(n+1)$ in the redundancy bound cannot be improved asymptotically. *Hint:* Apply the Hamming bound.

Problem 6: Reed–Muller codes and their duals

Let q be any prime power, $m \geq 1$, and $0 \leq r \leq m(q-1)$. A polynomial $f \in \mathbb{F}_q[X_1, \dots, X_m]$ is *reduced* if its degree in each variable is at most $q-1$. Define the Reed–Muller code

$$\text{RM}_q(r, m) = \left\{ (f(x))_{x \in \mathbb{F}_q^m} : f \text{ is reduced and } \deg f \leq r \right\} \subseteq \mathbb{F}_q^{q^m},$$

where $\deg f$ denotes total degree. For a linear code $\mathcal{C} \subseteq \mathbb{F}_q^N$, its dual is

$$\mathcal{C}^\perp = \{z \in \mathbb{F}_q^N : \sum_{i=1}^N z_i c_i = 0 \text{ for every } c \in \mathcal{C}\}.$$

- (a) **Minimum distance.** Prove the following common generalization of the distance lemmas shown in class. Suppose $r < m(q-1)$ and write $r = a(q-1) + b$, where $0 \leq b < q-1$. Then every nonzero reduced polynomial of total degree at most r is nonzero on at least

$$(q-b)q^{m-a-1}$$

points of $\mathbb{F}_q^m$. Exhibit a polynomial attaining this bound, and hence determine the exact minimum distance.

Hint: Induct on the number of variables, viewing f as a univariate polynomial whose coefficients are polynomials in the remaining variables. Apply the univariate root bound at points where its leading coefficient is nonzero.

- (b) **The dual is Reed–Muller.** For $0 \leq r < m(q-1)$, prove that

$$\text{RM}_q(r, m)^\perp = \text{RM}_q(m(q-1) - r - 1, m).$$

First prove orthogonality of the two codes, and then compare their dimensions. *Hint:* It suffices to check pairs of monomials. For monomials $\prod_i X_i^{a_i}$ and $\prod_i X_i^{b_i}$, the inner product of their codewords is

$$\sum_{x \in \mathbb{F}_q^m} \prod_{i=1}^m x_i^{a_i+b_i} = \prod_{i=1}^m \left(\sum_{t \in \mathbb{F}_q} t^{a_i+b_i} \right).$$

Use the degree bounds to show that at least one of these one-variable sums is zero. For the dimension comparison, consider the map $(e_i)_{i=1}^m \mapsto (q-1-e_i)_{i=1}^m$.