CS 225: Pseudorandomness	Prof. Salil Vadhan
Problem Set 5	
Assigned: Tue. Apr. 14, 2009	Due: Wed. Apr. 29, 2009(1 PM)

- Recall that your problem set solutions must be typed. You can email your solutions to `cs225-hw@eecs.harvard.edu`, or turn in it to MD138. You may write formulas or diagrams by hand. Aim for clarity and conciseness in your solutions, emphasizing the main ideas over low-level details.
- If you use LaTeX, please submit both the source (`.tex`) and the compiled file (`.ps`). Name your files `PS5-yourlastname`.
- Starred problems are extra credit.

Problem 6.2. (Concatenated Codes) For codes $\mathrm{Enc}_1 : \{1, \ldots, N\} \to \Sigma_1^{n_1}$ and $\mathrm{Enc}_2 : \Sigma_1 \to \Sigma_2^{n_2}$, their *concatenation* $\mathrm{Enc} : \{1, \ldots, N\} \to \Sigma_2^{n_1 n_2}$ is defined by

$$\mathrm{Enc}(m) = \mathrm{Enc}_2(\mathrm{Enc}_1(m)_1)\mathrm{Enc}_2(\mathrm{Enc}_1(m)_2) \cdots \mathrm{Enc}_2(\mathrm{Enc}_1(m)_{n_1}).$$

This is typically used as a tool for reducing alphabet size, e.g. with $\Sigma_2 = \{0, 1\}$.

1. Prove that if Enc_1 has minimum distance δ_1 and Enc_2 has minimum distance δ_2, then Enc has minimum distance at least $\delta_1\delta_2$.
2. Prove that if Enc_1 is $(1 - \varepsilon_1, \ell_1)$ list-decodable and Enc_2 is (δ_2, ℓ_2) list-decodable, then Enc is $((1 - \varepsilon_1\ell_2) \cdot \delta_2, \ell_1\ell_2)$ list-decodable.
3. By concatenating a Reed–Solomon code and a Hadamard code, show that for every $n \in \mathbb{N}$ and $\varepsilon > 0$, there is a (fully) explicit code $\mathrm{Enc} : \{0, 1\}^n \to \{0, 1\}^{\hat{n}}$ with blocklength $\hat{n} = O(n^2/\varepsilon^2)$ with minimum distance at least $1/2 - \varepsilon$. Furthermore, show that with blocklength $\hat{n} = \mathrm{poly}(n, 1/\varepsilon)$, we can obtain a code that is $(1/2-\varepsilon, \mathrm{poly}(1/\varepsilon))$ list-decodable in *polynomial time.* (Hint: the inner code can be decoded by brute force.)

Problem 6.3. (List Decoding implies Unique Decoding for Random Errors)

1. Suppose that $\mathcal{C} \subseteq \{0, 1\}^{\hat{n}}$ is a code with minimum distance at least $1/4$ and rate at most $\alpha\varepsilon^2$ for a fixed constant $\alpha > 0$ be determined below, and we transmit a codeword $c \in \mathcal{C}$ over a channel in which each bit is flipped with probability $1/2 - 2\varepsilon$. Show that if α is sufficiently small, then with all but exponentially small probability over the errors, c will be the unique codeword at distance at most $1/2 - \varepsilon$ from the received word r.
2. Using Problem 6.2, deduce that for every $\varepsilon > 0$ and $n \in \mathbb{N}$, there is an explicit code of blocklength $\hat{n} = \mathrm{poly}(n, 1/\varepsilon)$ that can be uniquely decoded from $(1/2 - 2\varepsilon)$ random errors as above in polynomial time.

3. Suppose that $\mathcal{C} \subseteq \Sigma^{\hat{n}}$ is a code with minimum distance at least $1-\varepsilon$, alphabet size $|\Sigma| = q \geq 1/(\alpha\varepsilon^2)$, and rate at most $\alpha\varepsilon$ for a fixed constant $\alpha > 0$ be determined below, and we transmit a codeword $c \in \mathcal{C}$ over a channel in which each symbol σ is replaced with a uniformly random symbol from $\Sigma \setminus \{\sigma\}$ with probability $1-3\varepsilon$. Show that if α is sufficiently small, then with all but exponentially small probability over the errors, c will be the unique codeword at distance at most $1-2\varepsilon$ from the received word r.

   Similar to Part 2, this implies that list-decoding algorithms for distance close to 1 yield unique decoding algorithms for random errors at noise rates close to 1.

Problem 6.4. (List-decoding Reed–Solomon Codes)

1. Show that there is a polynomial-time algorithm for list-decoding the Reed-Solomon codes of degree d over $\mathbb{F}_q$ up to distance $1-\sqrt{2d/q}$, improving the $1-2\sqrt{d/q}$ bound from lecture. (Hint: do not use fixed upper bounds on the individual degrees of the interpolating polynomial $Q(X,Y)$, but rather allow as many monomials as possible.)
2. (*) Improve the list-decoding radius further to $1-\sqrt{d/q}$ by using the 'multiple-roots' trick used in Section 6.2.4.

Problem 6.5. (Codes vs. Hashing) Given any code $\mathrm{Enc} : [N] \to [M]^{\hat{n}}$, we can obtain a family of hash functions $\mathcal{H} = \{h_i : [N] \to [M]\}_{i\in[\hat{n}]}$ defined by $h_i(x) = \mathrm{Enc}(x)_i$, and conversely.

1. Show that Enc has minimum distance at least δ iff $\mathcal{H}$ has collision probability at most $1-\delta$. That is, for all $x \neq y \in [N]$, we have $\Pr_i[h_i(x) = h_i(y)] \leq 1-\delta$. (This is a generalization of the definition of universal hash functions, which correspond to the case that $\delta = 1 - 1/M$.)
2. The Leftover Hash Lemma extends to families of functions with low collision probability; specifically if a family $\mathcal{H}$ with range $[M]$ has collision probability at most $(1+\varepsilon^2)/M$, then $\mathrm{Ext}(x,h) = (h, h(x))$ is a (k,ε) extractor for $k = m + 2\log(1/\varepsilon) + O(1)$, where $m = \log M$. Use this to prove the Johnson Bound for small alphabets: if a code $\mathrm{Enc} : [N] \to [M]^{\hat{n}}$ has minimum distance at least $1-1/M-\gamma/M$, then it is $(1-1/M-\sqrt{\gamma}, O(M/\gamma))$ list-decodable.

Problem 5. (Limitations on the Seed Length) Prove that a cryptographic pseudorandom generator cannot have seed length $\ell(n) = O(\log n)$. Note where your proof fails if we only require that it is an $(n^d, 1/n^d)$ pseudorandom generator for a fixed constant d.