Information and Coding Theory **Autumn 2014**

Homework 3

Due: December 4, 2014

Note: *You may discuss these problems in groups. However, you must write up your own solutions and mention the names of the people in your group. Also, please do mention any books, papers or other sources you refer to. It is recommended that you typeset your solutions in LaTeX.*

1. **More on linear codes.** Recall that a linear code $C : \mathbb{F}_q^k \to \mathbb{F}_q^n$ was specificied by a *generator matrix* G such that $\forall x \in \mathbb{F}_q^k,\ C(x) = Gx$. The *parity-check* matrix was a matrix H such that the columns of H^T form a basis for the null-space of G^T. Prove the following facts about linear codes.

   (a) Prove that for a linear code C, the distance $\Delta(C)$ can be written as

   $$\Delta(C) \ = \ \min_{z \in C \setminus \{0^n\}} \mathsf{wt}(z)\,,$$

   where 0^n denotes the all-zero vector in $\mathbb{F}_q^n$ and $\mathsf{wt}(z)$ denotes the number of non-zero entries in z.

   (b) Let $n = 2^r - 1$ for some integer r. Recall that the general Hamming code (over the field $\mathbb{F}_2$) is defined by the parity-check matrix $H \in \mathbb{F}_2^{r \times n}$ where the i^{th} column of H is given by the number i written in binary using r bits (take the top entry to be the most significant bit and the bottom entry to be the least significant bit). Find the message length, block length and the distance for this code.

   (c) For a linear code C with generator matrix G and parity-check matrix H, it's *dual code* $C^{\perp}$ is defined as a code with generator matrix H^T. Prove that G^T is a parity-check matrix for $C^{\perp}$. Find the message length, block length and distance for the *dual code* of the Hamming code defined above.

2. **Scrambled Reed-Solomon Codes [due to Venkat Guruswami].** Let $\{a_1, \ldots, a_n\}$ be distinct elements of $\mathbb{F}_q$ used to define a Reed-Solomon code $C : \mathbb{F}_q^k \to \mathbb{F}_q^n$. Assume that $k < n/6$. Recall that a message $(m_0, \ldots, m_{k-1})$ is encoded by thinking of it as a polynomial $P(X) = \sum_{j=0}^{k-1} m_j \cdot X^j$ and sending $(P(a_1), \ldots, P(a_n))$. For the following parts, assume the fact (used in class) that for a bivariate polynomial $Q(X,Y)$, we can find all its factors of the form $Y - f(X)$.

   (a) Suppose we sent two codewords according to the polynomials P and P' (of degree $k-1$) but they got mixed up. Thus, we now have two lists $(b_1, \ldots, b_n)$ and $(c_1, \ldots, c_n)$ and we know for each $i \in [n]$

   $$\text{either } P(a_i) = b_i \text{ and } P'(a_i) = c_i \quad \text{or} \quad P(a_i) = c_i \text{ and } P'(a_i) = b_i$$

Note that each coordinate could be independently scrambles i.e., it may happen that for some i, $P(a_i) = b_i$ and $P'(a_i) = c_i$ and for some $j \neq i$, $P(a_j) = c_j$ and $P'(a_j) = b_j$. Also, we don't know which is the case for which coordinate i.

Give an algorithm to find both P and P'. [**Hint**: First find $P + P'$ and $P \cdot P'$.]

(b) Now, suppose that instead of getting both the values $P(a_i)$ and $P'(a_i)$ for each i, we only got one value β_i, such that for each i we either have $\beta_i = P(a_i)$ or $\beta_i = P'(a_i)$. Again, it might happen that for some i, $\beta_i = P(a_i)$ while for some other $j \neq i$, $\beta_j = P'(a_j)$ and we don't know which is the case for which i. However, we are given the promise that

$$\frac{n}{3} \leq |\{i \in [n] \mid \beta_i = P(a_i)\}| \leq \frac{2n}{3} \quad \text{and} \quad \frac{n}{3} \leq \left|\{i \in [n] \mid \beta_i = P'(a_i)\}\right| \leq \frac{2n}{3}\,.$$

Give an algorithm to find both P and P'.