ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout XX Information Theory and Coding
Solutions to the Final Exam Jan. 24, 2013

PROBLEM 1.

(a) Since Y is a function of (X, K), $H(Y|X, K) = 0$. Similarly, $H(X|Y, K) = 0$.

(b) Since $Y = f(X, K)$, $H(Y|K) \leq H(X, K|K) = H(X|K)$. Similarly $H(X|K) \leq H(Y|K)$. Consequently the two conditional entropies are the same.

(c) We have $H(Y) \geq H(Y|K) = H(X|K) = H(X)$, the inequality is due to (b) and the last equality due to the independence of X and K.

(d) As Y is a function of (X, K), $H(Y|X) \leq H(X, K|X) = H(K|X) \leq H(K)$. Indeed, this is true even if X and K were dependent.

(e) From (c) we have $H(Y) \geq H(X)$, and from (d) we have $H(K) \geq H(Y|X)$. By independence of X and Y we have $H(Y|X) = H(Y)$, Consequently $H(K) \geq H(Y) \geq H(X)$.

(f) By assumption the system is secure for any distribution of X, so the inequality in (e) holds for any choice of p_X. Choosing p_X as the uniform distribution gives $H(K) \geq \log |\mathcal{X}|$.

(g) It is clear that g recovers X from $Y = f(X, K)$ and K. All that is needed to prove is that the system is secure. To that end, note that $H(Y|X) = H((X+K) \bmod m|X) = H(K|X) = H(K)$, where the last inequality is by the independence of X and K. Thus,

$$0 \leq I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(K) = H(Y) - \log m \leq \log m - \log m = 0.$$

(Note that we additionally obtain $H(Y) = \log m$.)

Moral of the story: information theoretically secure cryptosystems require a key with high entropy. This puts into doubt their practical feasibility: As an information theoretically secure system needs to have a mechanism to secretly communicate a key which has more entropy than the message itself, why not use this secure key-distribution mechanism to communicate the message?

PROBLEM 2.

(a) When x is sent it is impossible to receive a y for which $P(y|x) = 0$, consequently X and Y are compatible with probability 1. On the other hand, $\tilde{X}$ and Y will be incompatible whenever $Y \in \{0, 1\}$ and $\tilde{X} \neq Y$. The first event has probability $1 - p$ the second has probability $1/2$ and the two are independent. Thus, $\tilde{X}$ and Y are compatible with probability $1 - (1 - p)/2 = (1 + p)/2$.

(b) As in (a) X^n and Y^n are always compatible. For $\tilde{X}^n$, we need compatibility for each of the n indices, by independence this has probability $[(1 + p)/2]^n$.

(c) For the indices for which $Y_i = e$, the symbols $\tilde{X}_i$ and Y_i are always compatible. For the remaining indices $Y_i \in \{0, 1\}$, and $\tilde{X}_i$ and Y_i will be compatible with probability $1/2$. It follows that $\tilde{X}^n$ and Y^n are compatible with probability $(1/2)^{n-k}$ conditional on k erasures in Y^n.

(d) Since the transmitted message m and Y^n always satisfies the compatibility condition, the only way an error happens is when some other codeword also satisfies the compatibility condition (in which case the receiver declares 0.) By (b), for any $m' \neq m$, the probability that $X^n(m')$ and Y^n are compatible is $[(1 + p)/2]^n$. Since there are $M - 1$ choices of m', the union bound gives $P_e \leq (M - 1)[(1 + p)/2]^n$.

(e) The bound on the error probability above can we written as

$$P_e \leq 2^{n[R+\log[(1+p)/2]]}.$$

We thus conclude that as long as $R < \log(2/(1 + p)) =: R_0$, reliable communication is possible. (The quantity R_0 is known in the literature as the *cut-off rate*, and is also defined for arbitrary channels.)

(f) Let A denote the event that Y^n contains more than k erasures, and let B denote the complementary event, that Y^n contains k or fewer erasures. First write

$$P_e = \Pr\{\text{Error}|A\} \Pr\{A\} + \Pr\{\text{Error}|B\} \Pr\{B\}$$

Upper bound the $\Pr\{\text{Error}|A\}$ by 1, also upper bound $\Pr\{B\}$ by 1. We thus find

$$P_e \leq \Pr\{A\} + \Pr(\text{Error}|B\}.$$

By part (c), for any $m' \neq m$, the probability that $X^n(m')$ and Y^n will be compatible conditional on B is upper bounded by $(1/2)^{n-k}$. Union bound thus gives

$$\Pr(\text{Error}|B\} \leq (M - 1)\big(\tfrac{1}{2}\big)^{n-k}$$

and we obtain

$$P_e \leq \Pr(Y^n \text{ contains more than } k \text{ erasures}) + (M - 1)\big(\tfrac{1}{2}\big)^{n-k}.$$

(g) Fix $q > p$ and choose $k = nq$. By the law of large numbers the first term in the bound above approaches zero as n gets large. The second term also approaches zero if $R < 1 - q$. Thus, we conclude that for any $q > p$, it is possible to communicate reliably for all rates up to $1 - q$. Since q may be chosen as close to p as desired, we get $R_1 = 1 - p$. (Note that this is in fact the capacity of the BEC, and the conclusion on the rates we obtain is the best possible.)

PROBLEM 3.

(a) Since the input of the channel is (X_1, X_2) and the output if (A, Y), the mutual information between them is

$$I(X_1, X_2; A, Y) = I(X_1, X_2; A) + I(X_1, X_2; Y|A) = I(X_1, X_2; Y|A)$$

where the first equality is by the chain rule, second by the independence of A and the channel input.

(b) Given X_1, X_2, A, we know X_A. Thus $h(Y|X_1, X_2, A) = h(X_A + Z|X_1, X_2, A) = h(Z|X_1, X_2, A) = h(Z) = \frac{1}{2}\log(2\pi e)$. Note that we have used the independence of Z and X_1, X_2, A

(c) Conditional on $A = 1$, $Y = X_1 + Z$. Thus conditional on $A = 1$, the variance of Y is equal to $\mathrm{Var}(X_1) + \mathrm{Var}(Z) = \mathrm{Var}(X_1) + 1$. Recalling the bound on differential entropy in terms of variance, we have

$$h(Y|A = 1) \leq \tfrac{1}{2}\log(2\pi e(\mathrm{Var}(X_1) + 1)) \leq \tfrac{1}{2}\log(2\pi e(E[X_1^2] + 1)),$$

with equality if and only if X_1 is Gaussian and has zero mean.

(d) As $h(Y|A) = h(Y|A = 1)p + h(Y|A = 2)(1 - p)$, by (c) we see that $h(Y|A)$ is maximized with X_1 and X_2 are zero mean and Gaussian. It is neither necessary for them to be independent, not jointly Gaussian.

(e) By the previous parts, the mutual information between the input and the output is given by

$$p\tfrac{1}{2}\log(1 + P_1) + (1 - p)\tfrac{1}{2}\log(1 + P_2).$$

where $P_1 = E[X_1^2]$ and $P_2 = E[X_2^2]$. To find the capacity, we need to maximize the above over the choice of non-negative P_1, P_2 for which $P_1 + P_2 = 1$. Assuming optimizing (P_1, P_2) are both positive, the Kuhn-Tucker conditions give that for the optimal choice there is a λ for which

$$p = \lambda(1 + P_1) \quad \text{and} \quad 1 - p = \lambda(1 + P_2).$$

Summing the two equations give $1 = 3\lambda$, from which we obtain $P_1 = p - 1/3$, $P_2 = 2/3 - p$. For $p \in (1/3, 2/3)$ this gives the optimal choice. For $p \leq 1/3$ we need to choose $(P_1, P_2) = (0, 1)$, for $p \geq 2/3$ we need to choose $(P_1, P_2) = (1, 0)$.

What we studied here is a simplified case of 'diversity': the receiver receives one of two sent objects, but the transmitter does not know which it will be. It is intuitively clear that one should put more energy to the object that has a higher chance of reception; part (e) tells exactly how much. The "correlations don't matter" observation of (d) makes it possible to simplify design: X_1 and X_2 can be chosen as scaled versions of a common quantity X.

Problem 4.

(a) Any codeword of $\mathcal{C}$ is of the from $\langle \mathbf{a}, \mathbf{a} \oplus \mathbf{b} \rangle$ with $\mathbf{a} \in \mathcal{C}_1$ and $\mathbf{b} \in \mathcal{C}_2$. Given two codewords $\langle \mathbf{u}', \mathbf{u}' \oplus \mathbf{v}' \rangle$ and $\langle \mathbf{u}'', \mathbf{u}'' \oplus \mathbf{v}'' \rangle$ of $\mathcal{C}$, their sum is $\langle \mathbf{u}, \mathbf{u}+\mathbf{v} \rangle$ with $\mathbf{u} = \mathbf{u}' \oplus \mathbf{u}''$ and $\mathbf{v} = \mathbf{v}' \oplus \mathbf{v}''$. Since $\mathcal{C}_1$ and $\mathcal{C}_2$ are linear codes $\mathbf{u} \in \mathcal{C}_1$ and $\mathbf{v} \in \mathcal{C}_2$. Thus the sum of any two codewords of $\mathcal{C}$ is a codeword of $\mathcal{C}$ and we conclude that $\mathcal{C}$ is linear.

(b) If $(\mathbf{u}, \mathbf{v}) \neq (\mathbf{u}', \mathbf{v}')$, then either $\mathbf{u} \neq \mathbf{u}'$, or, $\mathbf{u} = \mathbf{u}'$ and $\mathbf{v} \neq \mathbf{v}'$. In either case $\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle \neq \langle \mathbf{u}' | \mathbf{u}' \oplus \mathbf{v}' \rangle$: in the first case the first halves differ, in the second case the second halves differ. Thus no two of the $(\mathbf{u}, \mathbf{v})$ pairs are mapped to the same element of $\mathcal{C}$, and the code has exactly $M_1 M_2$ elements. Its rate is $\frac{1}{2n} \log(M_1 M_2) = \frac{1}{2} R_1 + \frac{1}{2} R_2$.

(c) As $\mathbf{v} = \mathbf{u} \oplus \mathbf{u} \oplus \mathbf{v}$,

$$w_H(\mathbf{v}) = w_H(\mathbf{u} \oplus \mathbf{u} \oplus \mathbf{v}) \leq w_H(\mathbf{u}) + w_H(\mathbf{u} \oplus \mathbf{v})$$

by the triangle inequality. Noting that the right hand side is $w_H(\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle)$ completes the proof.

(d) If $\mathbf{v} = \mathbf{0}$ we have $\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle = \langle \mathbf{u} | \mathbf{u} \rangle$ which has twice the Hamming weight of $\mathbf{u}$. Otherwise (c) gives $w_H(\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle) \geq w_H(\mathbf{v})$.

(e) Since $\mathcal{C}$ is linear its minimum distance equals the minimum weight of its non-zero codewords. If $\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle$ is non-zero either $\mathbf{v} \neq \mathbf{0}$, or, $\mathbf{v} = \mathbf{0}$ and $\mathbf{u} \neq \mathbf{0}$. By (d), in the first case $w_H(\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle) \geq w_H(\mathbf{v}) \geq d_1$, in the second case $w_H(\langle \mathbf{u} | \mathbf{u} \oplus \mathbf{v} \rangle) \geq 2 w_H(\mathbf{u}) \geq 2 d_2$. Thus $d \geq \min\{2d_1, d_2\}$.

(f) Let $\mathbf{u}_0$ be the minimum weight non-zero codeword of $\mathcal{C}_1$ and let $\mathbf{v}_0$ be the minimum weight non-zero codeword of $\mathcal{C}_2$. Note that $\langle \mathbf{u}_0 | \mathbf{u}_0 \rangle$ is a non-zero codeword of $\mathcal{C}$ (corresponding to the choice $\mathbf{u} = \mathbf{u}_0$, $\mathbf{v} = \mathbf{0}$). It has weight $2d_1$. Similarly, $\langle \mathbf{0} | \mathbf{v}_0 \rangle$ is also a non-zero codeword of $\mathcal{C}$ (corresponding to the choice $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{v}_0$). It has weight d_2. Consequently $d \leq \min\{2d_1, d_2\}$. In light of (e) we find $d = \min\{2d_1, d_2\}$.

This method of constructing a longer code from two shorter ones is known under several names: 'Plotkin construction', 'bar product', '$(u|u+v)$ construction' appear regularly in the literature. Compare this method to the 'obvious' method of letting the codewords to be $\langle \mathbf{u} | \mathbf{v} \rangle$. The simple method has the same blocklength and rate as we have here, but its minimum distance is only $\min\{d_1, d_2\}$. The factor two gained in d_1 by the bar product is significant, and many practical code families can be built from very simple base codes by a recursive application of the bar product. Notable among them are the family of Reed–Muller codes.