ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 30 Information Theory and Coding

Solutions to Homework 12 Dec. 19, 2017

PROBLEM 1.

(a) Suppose $\mathbf{x}$ and $\mathbf{x}'$ are two codewords in $\mathcal{C}$. Then for $\forall i = 0, 1, \ldots, m-1$,

$$x_0 + x_1\alpha_i + \cdots + x_{n-1}\alpha_i^{n-1} = 0$$
$$x'_0 + x'_1\alpha_i + \cdots + x'_{n-1}\alpha_i^{n-1} = 0$$

Therefore,

$$(x_0 + x'_0) + (x_1 + x'_1)\alpha_i + \cdots + (x_{n-1} + x'_{n-1})\alpha_i^{n-1} = 0 \qquad \text{for } \forall i = 0, 1, \ldots, m-1.$$

which shows $\mathbf{x} + \mathbf{x}'$ is also a codeword.

(b) $x(D) = x_0 + x_1 D + \cdots + x_{n-1}D^{n-1}$ is a polynomial of degree (at most) $n-1$ and $(x_0, \ldots, x_{n-1})$ is a codeword if $\alpha_0, \alpha_1, \ldots, \alpha_{m-1}$ are m of its roots. This means

$$x(D) = (D - \alpha_0)(D - \alpha_1)\ldots(D - \alpha_{m-1})h(D) = g(D)h(D)$$

for some $h(D)$. Note that $h(D)$ can have degree (at most) $n - m - 1$. On the other side, there is a one-to-one correspondence between the codewords of $\mathcal{C}$ and degree $n-1$ polynomials. Since $g(D)$ is fixed for all codewords, a polynomial $x(D)$ corresponding to a codeword $\mathbf{x}$ is determined by choosing the coefficients of $h(D) = h_0 + h_1 D + \cdots + h_{n-m-1}D^{n-m-1}$. Since $h_j \in \mathcal{X}$ for $j = 0, 1, \ldots, n-m-1$ we have q^{n-m} different $h(D)$s and, thus, q^{n-m} codewords.

(c) For every column vector $\mathbf{u} = [u_0, u_1, \ldots, u_{m-1}]^T$, $A\mathbf{u} = [u(1), u(\beta), \ldots, u(\beta^{n-1})]^T$. Consequently, $\mathbf{u}A = \mathbf{0}$ means $u(D)$ has n roots which is impossible (since it is a polynomial of degree $m - 1 < n$).

(d) Using the same reasoning as in (c) one can verify that $\mathbf{x} = (x_1, \ldots, x_n)$ is a codeword iff $\mathbf{x}A = \mathbf{0}$. This means A is the parity-check matrix of the code $\mathcal{C}$. Since the code is linear, using Problem 4 of Homework 11 we know that has minimum distance d iff every $d-1$ rows of H are linearly independent and some d rows are linearly dependent. That A has rank m implies there are no m linearly dependent rows thus $d \geq m+1$. On the other side, we know from the Singleton bound that a code with q^{n-m} codewords and block-length n has minimum distance $d \leq m+1$. Thus we conclude that $d = m+1$.

PROBLEM 2.

(a) For every $0 \leq p \leq 1$, define $\overline{p} := 1 - p$. We have:

$$h_2(\overline{p}) = -\overline{p}\log\overline{p} - p\log p = -p\log p - \overline{p}\log\overline{p} = h_2(p). \qquad (1)$$

On the other hand, it is easy to check that for every $0 \leq p', p'' \leq 1$, we have:

$$\overline{p'} * p'' = p' * \overline{p''} = \overline{p' * p''} \quad \text{and} \quad \overline{p'} * \overline{p''} = p' * p''.$$

Now (1) implies that

$$h_2(\overline{p'} * p'') = h_2(p' * \overline{p''}) = h_2(\overline{p'} * \overline{p''}) = h_2(p' * p''). \qquad (2)$$

Let $p' = \mathbb{P}[X_1 = 1]$ and $p'' = \mathbb{P}[X_2 = 1]$. We have the following:

- $\mathbb{P}[X_1 \oplus X_2 = 1] = \mathbb{P}[X_1 = 1]\mathbb{P}[X_2 = 0] + \mathbb{P}[X_1 = 0]\mathbb{P}[X_2 = 1] = p'\overline{p''} + \overline{p'}p'' = p' * p''$. Therefore, $H(X_1 \oplus X_2) = h_2(p' * p'')$.
- Since $H(X_1) = h_2(p_1)$, then we have either $p' = p_1$ or $p' = 1 - p_1$. I.e., we have $p_1 = p'$ or $p_1 = 1 - p' = \overline{p'}$.
- Since $H(X_2) = h_2(p_2)$, then we have either $p'' = p_2$ or $p'' = 1 - p_2$. I.e., we have $p_2 = p''$ or $p_2 = 1 - p'' = \overline{p''}$.

Now (2) implies that $H(X_1 \oplus X_2) = h_2(p' * p'') = h_2(p_1 * p_2)$.

(b) We have $H(X_1|Y) = \sum_{y\in\mathcal{Y}} H(X_1|Y = y)\mathbb{P}_Y(y) = \sum_{y\in\mathcal{Y}} h_2(p_1(y))q(y)$.

Now for every $y \in \mathcal{Y}$, X_1 and X_2 are independent conditioned on $Y = y$. Moreover, $H(X_1|Y = y) = h_2(p_1(y))$ and $H(X_2|Y = y) = H(X_2) = h_2(p_2)$ since X_2 and Y are independent. Therefore, part (a) implies that $H(X_1 \oplus X_2|Y = y) = h_2(p_1(y) * p_2)$.

We conclude that

$$\begin{aligned} H(X_1 \oplus X_2|Y) &= \sum_{y\in\mathcal{Y}} H(X_1 \oplus X_2|Y = y)\mathbb{P}_Y(y) \\ &= \sum_{y\in\mathcal{Y}} h_2(p_1(y) * p_2)q(y) = \sum_{y\in\mathcal{Y}} h_2(p_2 * p_1(y))q(y). \end{aligned}$$

(c) Note that $p_2 * p = p(1 - p_2) + p_2(1 - p) = \beta p + p_2$, where $\beta = 1 - 2p_2 \geq 0$. Let $g(p) = \frac{\frac{\partial}{\partial p}h_2(p_2 * p)}{\frac{\partial}{\partial p}h_2(p)} = \frac{\frac{\partial}{\partial p}h_2(\beta p + p_2)}{\frac{\partial}{\partial p}h_2(p)} = \frac{\beta h_2'(\beta p+p_2)}{h_2'(p)}$. We have

$$\begin{aligned} g'(p) &= \frac{\beta^2 h_2''(\beta p + p_2)h_2'(p) - \beta h_2''(p)h_2'(\beta p + p_2)}{h_2'(p)^2} \\ &= \frac{\beta h_2''(\beta p + p_2)h_2''(p)}{h_2'(p)^2}\left[\beta\frac{h_2'(p)}{h_2''(p)} - \frac{h_2'(\beta p + p_2)}{h_2''(\beta p + p_2)}\right]. \end{aligned}$$

Note that $h_2'(p) = \log\frac{1-p}{p}$ and $h_2''(p) = \frac{-1}{p(1-p)\ln 2}$, which implies that $h_2''(\beta p + p_2) \leq 0$ and $h_2''(p) \leq 0$. Therefore, $\frac{\beta h_2''(\beta p+p_2)h_2''(p)}{h_2'(p)^2} \geq 0$ and so it is sufficient to show that we have $\beta\frac{h_2'(p)}{h_2''(p)} - \frac{h_2'(\beta p+p_2)}{h_2''(\beta p+p_2)} \geq 0$. Now define $\alpha = 1 - 2p$. It is easy to check the following:

- $p = \frac{1}{2}(1 - \alpha)$.
- $1 - p = \frac{1}{2}(1 + \alpha)$.
- $\beta p + p_2 = \frac{1}{2}(1 - \alpha\beta)$.
- $1 - (\beta p + p_2) = \frac{1}{2}(1 + \alpha\beta)$.

Therefore, we have

$$\beta\frac{h_2'(p)}{h_2''(p)} = -\beta(\ln 2)p(1 - p)\log\frac{1 - p}{p} = -\frac{\beta\ln 2}{4}(1 - \alpha^2)\log\frac{1 + \alpha}{1 - \alpha},$$

and

$$\frac{h_2'(\beta p + p_2)}{h_2''(\beta p + p_2)} = -(\ln 2)(\beta p+p_2)(1-\beta p-p_2)\log\frac{1 - \beta p - p_2}{\beta p + p_2} = -\frac{\ln 2}{4}(1-(\alpha\beta)^2)\log\frac{1 + \alpha\beta}{1 - \alpha\beta}.$$

Using the formula $\log(1+x) = \sum_{k\geq 1}(-1)^{k-1}\frac{x^k}{k}$, we get

$$\begin{aligned}
\log\frac{1+x}{1-x} = \log(1+x) - \log(1-x) &= \left(\sum_{k\geq 1}(-1)^{k-1}\frac{x^k}{k}\right) - \left(\sum_{k\geq 1}(-1)^{k-1}\frac{(-x)^k}{k}\right) \\
&= \sum_{k\geq 1}\left((-1)^{k-1}+1\right)\frac{x^k}{k} = 2\sum_{\substack{k\geq 1\\ k \text{ is odd}}}\frac{x^k}{k}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
-(1-x^2)\log\frac{1+x}{1-x} &= -2\sum_{\substack{k\geq 1\\ k \text{ is odd}}}\frac{x^k}{k} + 2\sum_{\substack{k\geq 1\\ k \text{ is odd}}}\frac{x^{k+2}}{k} = -2x - 2\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\frac{x^k}{k} + 2\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\frac{x^k}{k-2} \\
&= -2x + 2\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)x^k.
\end{aligned}$$

Hence,

$$\begin{aligned}
\beta\frac{h_2'(p)}{h_2''(p)} &= -\frac{\beta\ln 2}{4}(1-\alpha^2)\log\frac{1+\alpha}{1-\alpha} = \frac{\beta\ln 2}{4}\left[-2\alpha + 2\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)\alpha^k\right] \\
&= -\frac{\alpha\beta\ln 2}{2} + \frac{\ln 2}{2}\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)\beta\alpha^k,
\end{aligned}$$

and

$$\begin{aligned}
\frac{h_2'(\beta p + p_2)}{h_2''(\beta p + p_2)} &= -\frac{\ln 2}{4}(1-(\alpha\beta)^2)\log\frac{1+\alpha\beta}{1-\alpha\beta} = \frac{\ln 2}{4}\left[-2\alpha\beta + 2\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)(\alpha\beta)^k\right] \\
&= -\frac{\alpha\beta\ln 2}{2} + \frac{\ln 2}{2}\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)\beta^k\alpha^k.
\end{aligned}$$

We conclude that

$$\beta\frac{h_2'(p)}{h_2''(p)} - \frac{h_2'(\beta p + p_2)}{h_2''(\beta p + p_2)} = \frac{\ln 2}{2}\sum_{\substack{k\geq 3\\ k \text{ is odd}}}\left(\frac{1}{k-2}-\frac{1}{k}\right)(\beta-\beta^k)\alpha^k \overset{(*)}{\geq} 0,$$

where $(*)$ follows from the fact that $\beta = 1-2p_2 \leq 1$ which implies that $\beta^k \leq \beta$. Therefore, $g'(p) \geq 0$ and so $g(p)$ is increasing. We conclude that the function f is convex.

(d) We have

$$\begin{aligned}
H(X_1 \oplus X_2|Y) &= \sum_{y\in\mathcal{Y}} h_2(p_2 * p_1(y))q(y) = \sum_{y\in\mathcal{Y}} h_2\big(p_2 * h_2^{-1}\big(H(X_1|Y=y)\big)\big)q(y) \\
&= \sum_{y\in\mathcal{Y}} f\big(H(X_1|Y=y)\big)q(y) \overset{(*)}{\geq} f\Big(\sum_{y\in\mathcal{Y}} H(X_1|Y=y)q(y)\Big) \\
&= f(H(X_1|Y)) = h_2\big(p_2 * h_2^{-1}\big(H(X_1|Y)\big)\big) = h_2(p_2 * p_1) = h_2(p_1 * p_2),
\end{aligned}$$

where $(*)$ follows from the convexity of the function f.

(e) For every $y_1 \in \mathcal{Y}_1$, let $0 \le p_1(y_1) \le \frac{1}{2}$ be such that $H(X_1|Y_1 = y_1) = h_2(p_1(y_1))$ and let $q_1(y_1) = \mathbb{P}_{Y_1}(y_1)$. Similarly, for every $y_2 \in \mathcal{Y}_2$, let $0 \le p_2(y_2) \le \frac{1}{2}$ be such that $H(X_2|Y_2 = y_2) = h_2(p_2(y_2))$ and let $q_2(y_2) = \mathbb{P}_{Y_2}(y_2)$. For every $y_1 \in \mathcal{Y}_1$, define the mapping $f_{y_1} : [0,1] \to \mathbb{R}$ as $f_{y_1}(h) = h_2(p_1(y) * h_2^{-1}(h))$. Part (c) implies that f_{y_1} is convex for every $y_1 \in \mathcal{Y}_1$. We have

$$
\begin{aligned}
H(X_1 \oplus X_2|Y_1, Y_2) &= \sum_{y_1\in\mathcal{Y}_1}\sum_{y_2\in\mathcal{Y}_2} h_2(p_1(y_1) * p_2(y_2))\mathbb{P}_{Y_1,Y_2}(y_1, y_2) \\
&= \sum_{y_1\in\mathcal{Y}_1}\sum_{y_2\in\mathcal{Y}_2} h_2(p_1(y_1) * p_2(y_2))q_1(y_1)q_2(y_2) \\
&= \sum_{y_1\in\mathcal{Y}_1} q_1(y_1) \sum_{y_2\in\mathcal{Y}_2} h_2\big(p_1(y_1) * h_2^{-1}\big(H(X_2|Y_2 = y_2)\big)\big)q_2(y_2) \\
&= \sum_{y_1\in\mathcal{Y}_1} q_1(y_1) \sum_{y_2\in\mathcal{Y}_2} f_{y_1}\big(H(X_2|Y_2 = y_2)\big)q_2(y_2) \\
&\overset{(*)}{\ge} \sum_{y_1\in\mathcal{Y}_1} q_1(y_1)f_{y_1}\Big(\sum_{y_2\in\mathcal{Y}_2} H(X_2|Y_2 = y_2)q_2(y_2)\Big) \\
&= \sum_{y_1\in\mathcal{Y}_1} q_1(y_1)f_{y_1}(H(X_2|Y_2)) = \sum_{y_1\in\mathcal{Y}_1} q_1(y_1)h_2\big(p_1(y_1) * h_2^{-1}\big(H(X_2|Y_2)\big)\big) \\
&= \sum_{y_1\in\mathcal{Y}_1} q_1(y_1)h_2(p_1(y_1) * p_2) = \sum_{y_1\in\mathcal{Y}_1} h_2\big(p_2 * h_2^{-1}\big(H(X_1|Y_1 = y_1)\big)\big)q_1(y_1) \\
&= \sum_{y_1\in\mathcal{Y}_1} f\big(H(X_1|Y_1 = y_1)\big)q_1(y_1) \overset{(**)}{\ge} f\Big(\sum_{y_1\in\mathcal{Y}_1} H(X_1|Y_1 = y_1)q(y_1)\Big) \\
&= f(H(X_1|Y_1)) = h_2\big(p_2 * h_2^{-1}\big(H(X_1|Y_1)\big)\big) = h_2(p_2 * p_1) = h_2(p_1 * p_2),
\end{aligned}
$$

where $(*)$ follows from the convexity of the functions $\{f_{y_1} : \ y_1 \in \mathcal{Y}_1\}$ and $(**)$ follows from the convexity of f.

PROBLEM 3.

(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 $\langle \mathbf{u}, \mathbf{u} \oplus \mathbf{v}\rangle$ 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 $(\mathbf{u}, \mathbf{v}) \ne (\mathbf{u}', \mathbf{v}')$, then either $\mathbf{u} \ne \mathbf{u}'$, or, $\mathbf{u} = \mathbf{u}'$ and $\mathbf{v} \ne \mathbf{v}'$. In either case $\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v}\rangle \ne \langle \mathbf{u}'|\mathbf{u}' \oplus \mathbf{v}'\rangle$: 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_1M_2 elements. Its rate is $\frac{1}{2n}\log(M_1M_2) = \frac{1}{2}R_1 + \frac{1}{2}R_2$.

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

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

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 $w_H(\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v}\rangle) \ge w_H(\mathbf{v})$.

(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 $w_H(\langle \mathbf{u}|\mathbf{u} \oplus \mathbf{v}\rangle) \geq 2w_H(\mathbf{u}) \geq 2d_2$. 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 block-length 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.

Problem 4.

(a) We have

$$\begin{aligned}
W^-(y_1, y_2|u_1) &= \mathbb{P}_{Y_1,Y_2|X_1\oplus X_2}(y_1, y_2|u_1) = \frac{\mathbb{P}_{Y_1,Y_2,X_1\oplus X_2}(y_1, y_2, u_1)}{\mathbb{P}_{X_1\oplus X_2}(u_1)} \\
&\overset{(*)}{=} 2\mathbb{P}_{Y_1,Y_2,X_1\oplus X_2}(y_1, y_2, u_1) \\
&= 2\sum_{u_2\in\{0,1\}} \mathbb{P}_{Y_1,Y_2,X_1\oplus X_2,X_2}(y_1, y_2, u_1, u_2) \\
&\overset{(**)}{=} 2\sum_{u_2\in\{0,1\}} \mathbb{P}_{Y_1,Y_2,X_1,X_2}(y_1, y_2, u_1 \oplus u_2, u_2) \\
&= 2\sum_{u_2\in\{0,1\}} \mathbb{P}_{Y_1,Y_2|X_1,X_2}(y_1, y_2|u_1 \oplus u_2, u_2)\mathbb{P}_{X_1,X_2}(u_1 \oplus u_2, u_2) \\
&= 2\sum_{u_2\in\{0,1\}} W(y_1|u_1 \oplus u_2)W(y_2|u_2)\frac{1}{2^2} \\
&= \frac{1}{2}\sum_{u_2\in\{0,1\}} W(y_1|u_1 \oplus u_2)W(y_2|u_2),
\end{aligned}$$

where $(*)$ follows from the fact that if X_1, X_2 are independent and uniform then $X_1 \oplus X_2$ is also uniform. $(**)$ follows from the fact that

$$(X_1 \oplus X_2 = u_1 \text{ and } X_2 = u_2) \Leftrightarrow (X_1 = u_1 \oplus u_2 \text{ and } X_2 = u_2).$$

(b) We have

$$
\begin{aligned}
W^+(y_1,y_2,u_1|u_2) &= \mathbb{P}_{Y_1,Y_2,X_1\oplus X_2|X_2}(y_1,y_2,u_1|u_2) = \frac{\mathbb{P}_{Y_1,Y_2,X_1\oplus X_2,X_2}(y_1,y_2,u_1,u_2)}{\mathbb{P}_{X_2}(u_2)}\\
&= 2\mathbb{P}_{Y_1,Y_2,X_1\oplus X_2,X_2}(y_1,y_2,u_1,u_2)\\
&\overset{(*)}{=} 2\mathbb{P}_{Y_1,Y_2,X_1,X_2}(y_1,y_2,u_1\oplus u_2,u_2)\\
&= 2\mathbb{P}_{Y_1,Y_2|X_1,X_2}(y_1,y_2|u_1\oplus u_2,u_2)\mathbb{P}_{X_1,X_2}(u_1\oplus u_2,u_2)\\
&= 2W(y_1|u_1\oplus u_2)W(y_2|u_2)\frac{1}{2^2}\\
&= \frac{1}{2}W(y_1|u_1\oplus u_2)W(y_2|u_2),
\end{aligned}
$$

where $(*)$ follows from the fact that

$$(X_1\oplus X_2 = u_1 \text{ and } X_2=u_2) \Leftrightarrow (X_1 = u_1\oplus u_2 \text{ and } X_2 = u_2).$$

(c) We have

$$
\begin{aligned}
Z(W^+) &= \sum_{\substack{y_1,y_2\in\mathcal{Y},\\ u_1\in\{0,1\}}} \sqrt{W^+(y_1,y_2,u_1|0)W^+(y_1,y_2,u_1|1)}\\
&=\frac{1}{2}\sum_{\substack{y_1,y_2\in\mathcal{Y},\\ u_1\in\{0,1\}}} \sqrt{W(y_1|u_1\oplus 0)W(y_2|0)W(y_1|u_1\oplus 1)W(y_2|1)}\\
&=\frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|0\oplus 0)W(y_2|0)W(y_1|0\oplus 1)W(y_2|1)}\right)\\
&\quad+\frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|1\oplus 0)W(y_2|0)W(y_1|1\oplus 1)W(y_2|1)}\right)\\
&=\frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|0)W(y_2|0)W(y_1|1)W(y_2|1)}\right)\\
&\quad+\frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W(y_1|1)W(y_2|0)W(y_1|0)W(y_2|1)}\right)\\
&=\frac{1}{2}\left(\sum_{y_1\in\mathcal{Y}} \sqrt{W(y_1|0)W(y_1|1)}\right)\left(\sum_{y_2\in\mathcal{Y}} \sqrt{W(y_2|0)W(y_2|1)}\right)\\
&\quad+\frac{1}{2}\left(\sum_{y_1\in\mathcal{Y}} \sqrt{W(y_1|0)W(y_1|1)}\right)\left(\sum_{y_2\in\mathcal{Y}} \sqrt{W(y_2|0)W(y_2|1)}\right)\\
&=\frac{1}{2}Z(W)\cdot Z(W) + \frac{1}{2}Z(W)\cdot Z(W) = Z(W)^2.
\end{aligned}
$$

(d) For every $y_1,y_2\in\mathcal{Y}$, we have:

$$
\begin{aligned}
W^-(y_1,y_2|0) &= \frac{1}{2}\sum_{u_2\in\{0,1\}} W(y_1|0\oplus u_2)W(y_2|u_2) = \frac{1}{2}\sum_{u_2\in\{0,1\}} W(y_1|u_2)W(y_2|u_2)\\
&= \frac{1}{2}W(y_1|0)W(y_2|0) + \frac{1}{2}W(y_1|1)W(y_2|1) = \frac{1}{2}\alpha(y_1)\alpha(y_2) + \frac{1}{2}\beta(y_1)\beta(y_2)\\
&= \frac{1}{2}(\alpha(y_1)\alpha(y_2)+\beta(y_1)\beta(y_2)),
\end{aligned}
$$

and

$$\begin{aligned}
W^-(y_1, y_2|1) &= \frac{1}{2}\sum_{u_2\in\{0,1\}} W(y_1|1\oplus u_2)W(y_2|u_2)\\
&= \frac{1}{2}W(y_1|1\oplus 0)W(y_2|0) + \frac{1}{2}W(y_1|1\oplus 1)W(y_2|1)\\
&= \frac{1}{2}W(y_1|1)W(y_2|0) + \frac{1}{2}W(y_1|0)W(y_2|1) = \frac{1}{2}\beta(y_1)\alpha(y_2) + \frac{1}{2}\alpha(y_1)\beta(y_2)\\
&= \frac{1}{2}(\alpha(y_1)\beta(y_2) + \beta(y_1)\alpha(y_2)).
\end{aligned}$$

We have

$$\begin{aligned}
Z(W^-) &= \sum_{y_1,y_2\in\mathcal{Y}} \sqrt{W^-(y_1,y_2|0)W^-(y_1,y_2|1)}\\
&= \frac{1}{2}\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{\Big(\alpha(y_1)\alpha(y_2) + \beta(y_1)\beta(y_2)\Big)\Big(\alpha(y_1)\beta(y_2) + \beta(y_1)\alpha(y_2)\Big)}.
\end{aligned}$$

(e) For every $x, y \geq 0$, we have $x + y \leq x + y + 2\sqrt{xy} = (\sqrt{x} + \sqrt{y})^2$ which implies that $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$. Therefore, for every $x, y, z, t \geq 0$ we have:

$$\sqrt{x+y+z+t} \leq \sqrt{x+y} + \sqrt{z+t} \leq \sqrt{x} + \sqrt{y} + \sqrt{z} + \sqrt{t}.$$

Therefore,

$$\begin{aligned}
&Z(W^-)\\
&\quad= \frac{1}{2}\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{\Big(\alpha(y_1)\alpha(y_2) + \beta(y_1)\beta(y_2)\Big)\Big(\alpha(y_1)\beta(y_2) + \beta(y_1)\alpha(y_2)\Big)}\\
&\quad= \frac{1}{2}\sum_{y_1,y_2\in\mathcal{Y}} \sqrt{\alpha(y_1)^2\gamma(y_2)^2 + \alpha(y_2)^2\gamma(y_1)^2 + \beta(y_2)^2\gamma(y_1)^2 + \beta(y_1)^2\gamma(y_2)^2}\\
&\quad\overset{(*)}{\leq} \frac{1}{2}\sum_{y_1,y_2\in\mathcal{Y}} \left(\sqrt{\alpha(y_1)^2\gamma(y_2)^2} + \sqrt{\alpha(y_2)^2\gamma(y_1)^2} + \sqrt{\beta(y_2)^2\gamma(y_1)^2} + \sqrt{\beta(y_1)^2\gamma(y_2)^2}\right)\\
&\quad= \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \alpha(y_1)\gamma(y_2)\right) + \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \alpha(y_2)\gamma(y_1)\right)\\
&\qquad+ \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \beta(y_2)\gamma(y_1)\right) + \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \beta(y_1)\gamma(y_2)\right),
\end{aligned}$$

where $(*)$ follows from the inequality $\sqrt{x+y+z+t} \leq \sqrt{x} + \sqrt{y} + \sqrt{z} + \sqrt{t}$.

(f) Note that $\sum_{y\in\mathcal{Y}} \alpha(y) = \sum_{y\in\mathcal{Y}} \beta(y) = 1$ and $\sum_{y\in\mathcal{Y}} \gamma(y) = Z(W)$. Therefore,

$$\begin{aligned}
Z(W^-) &\le \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \alpha(y_1)\gamma(y_2)\right) + \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \alpha(y_2)\gamma(y_1)\right) \\
&\quad + \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \beta(y_2)\gamma(y_1)\right) + \frac{1}{2}\left(\sum_{y_1,y_2\in\mathcal{Y}} \beta(y_1)\gamma(y_2)\right) \\
&= \frac{1}{2}\left(\sum_{y_1\in\mathcal{Y}} \alpha(y_1)\right)\left(\sum_{y_2\in\mathcal{Y}} \gamma(y_2)\right) + \frac{1}{2}\left(\sum_{y_2\in\mathcal{Y}} \alpha(y_2)\right)\left(\sum_{y_1\in\mathcal{Y}} \gamma(y_1)\right) \\
&\quad + \frac{1}{2}\left(\sum_{y_2\in\mathcal{Y}} \beta(y_2)\right)\left(\sum_{y_1\in\mathcal{Y}} \gamma(y_1)\right) + \frac{1}{2}\left(\sum_{y_1\in\mathcal{Y}} \beta(y_1)\right)\left(\sum_{y_2\in\mathcal{Y}} \gamma(y_2)\right) \\
&= \frac{1}{2}1\cdot Z(W) + \frac{1}{2}1\cdot Z(W) + \frac{1}{2}1\cdot Z(W) + \frac{1}{2}1\cdot Z(W) = 2Z(W).
\end{aligned}$$

Problem 5.

(a) We have

$$\begin{aligned}
Q_{i+1} = \sqrt{Z_{i+1}(1-Z_{i+1})} &= \begin{cases} \sqrt{Z_i^2(1-Z_i^2)} & \text{w.p. } 1/2 \\ \sqrt{(2Z_i - Z_i^2)(1-2Z_i+Z_i^2)} & \text{w.p. } 1/2 \end{cases} \\
&= \begin{cases} \sqrt{Z_i^2(1-Z_i)(1+Z_i)} & \text{w.p. } 1/2 \\ \sqrt{(2-Z_i)Z_i(1-Z_i)^2} & \text{w.p. } 1/2 \end{cases} \\
&= \begin{cases} \sqrt{Z_i(1-Z_i)}\sqrt{Z_i(1+Z_i)} & \text{w.p. } 1/2 \\ \sqrt{Z_i(1-Z_i)}\sqrt{(2-Z_i)(1-Z_i)} & \text{w.p. } 1/2 \end{cases} \\
&= \sqrt{Z_i(1-Z_i)}\begin{cases} \sqrt{Z_i(1+Z_i)} & \text{w.p. } 1/2 \\ \sqrt{(2-Z_i)(1-Z_i)} & \text{w.p. } 1/2 \end{cases} \\
&= Q_i\begin{cases} f_1(Z_i) & \text{w.p. } 1/2 \\ f_2(Z_i) & \text{w.p. } 1/2 \end{cases},
\end{aligned}$$

where $f_1(z) = \sqrt{z(z+1)}$ and $f_2(z) = \sqrt{(2-z)(1-z)}$.

(b) We have

$$f_1'(z) = \frac{2z+1}{2\sqrt{z(z+1)}}$$

so

$$\begin{aligned}
f_1''(z) &= \frac{4\sqrt{z(z+1)} - (2z+1)\dfrac{2(2z+1)}{2\sqrt{z(z+1)}}}{\left(2\sqrt{z(z+1)}\right)^2} \\
&= \frac{4z(z+1)-(2z+1)^2}{4\big(z(z+1)\big)^{\frac{3}{2}}} = \frac{-1}{4\big(z(z+1)\big)^{\frac{3}{2}}} \le 0.
\end{aligned}$$

Therefore, f_1 is concave. By noticing that $f_2(z) = f_1(1-z)$, we obtain:

$$\begin{aligned} f_1(z) + f_2(z) = f_1(z) + f_1(1-z) &= 2\left(\frac{1}{2}f_1(z) + \frac{1}{2}f_1(1-z)\right) \\ &\overset{(*)}{\leq} 2f_1\left(\frac{1}{2}z + \frac{1}{2}(1-z)\right) = 2f_1\left(\frac{1}{2}\right) = 2\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)} \\ &= 2\sqrt{\frac{1}{2}\cdot\frac{3}{2}} = 2\frac{\sqrt{3}}{2} = \sqrt{3}, \end{aligned}$$

where $(*)$ follows from the concavity of f_1. We have

$$\mathbb{E}\big[Q_{i+1} \mid Z_0, \ldots, Z_i\big] = \frac{1}{2}f_1(Z_i)Q_i + \frac{1}{2}f_2(Z_i)Q_i = \frac{1}{2}(f_1(Z_i) + f_2(Z_i))Q_i \leq \rho Q_i,$$

where $\rho = \frac{\sqrt{3}}{2} < 1$.

(c) We will show the claim by induction on $i \geq 0$. For $i = 0$, we have $Z_0 = z_0$ with probability 1. Therefore, $\mathbb{E}Q_0 = \sqrt{z_0(1-z_0)}$.

It is easy to that the function $[0,1] \to \mathbb{R}$ defined by $z \to \sqrt{z(1-z)}$ achieves its maximum at $z = \frac{1}{2}$, and so $\mathbb{E}Q_0 = \sqrt{z_0(1-z_0)} \leq \sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)} = \frac{1}{2}$. Therefore, the claim is true for $i = 0$.

Now suppose that the claim is true for $i \geq 0$, i.e., $\mathbb{E}Q_i \leq \frac{1}{2}\rho^i$. We have

$$\mathbb{E}Q_{i+1} = \mathbb{E}\left[\mathbb{E}\big[Q_{i+1} \mid Z_0, \ldots, Z_i\big]\right] \overset{(*)}{\leq} \mathbb{E}[\rho Q_i] = \rho\mathbb{E}[Q_i] \overset{(**)}{\leq} \rho \cdot \frac{1}{2}\rho^i = \frac{1}{2}\rho^{i+1},$$

where $(*)$ follows from Part (b) and $(**)$ follows from the induction hypothesis. We conclude that $\mathbb{E}Q_i \leq \frac{1}{2}\rho^i$ for every $i \geq 0$.

(d) By noticing that $\delta < z < 1-\delta$ if and only if $z(1-z) > \delta(1-\delta)$, we get:

$$\begin{aligned} \mathbb{P}\big[Z_i \in (\delta, 1-\delta)\big] &= \mathbb{P}\big[Z_i(1-Z_i) > \delta(1-\delta)\big] = \mathbb{P}\big[\sqrt{Z_i(1-Z_i)} > \sqrt{\delta(1-\delta)}\big] \\ &= \mathbb{P}\big[Q_i > \sqrt{\delta(1-\delta)}\big] \overset{(*)}{\leq} \frac{\mathbb{E}Q_i}{\sqrt{\delta(1-\delta)}} \overset{(**)}{\leq} \frac{\rho^i}{2\sqrt{\delta(1-\delta)}}, \end{aligned}$$

where $(*)$ follows from the Markov inequality and $(**)$ follows from Part (c). Now since $\rho < 1$, we have $\dfrac{\rho^i}{2\sqrt{\delta(1-\delta)}} \to 0$ as $i \to \infty$. We conclude that

$$\mathbb{P}\big[Z_i \in (\delta, 1-\delta)\big] \to 0 \text{ as } i \text{ gets large.}$$