PROBLEM 1. In a hypothesis testing problem with hypothesis $H \in \{0, 1, 2\}$, observation $Y \in \mathcal{Y}$, decision $\hat{H} \in \{0, 1, 2\}$, the penalty for deciding j when the true hypothesis is i is $\text{pen}(i, j) = |i - j|$.

(a) (3 pts) Suppose D_0, D_1 and D_2 are the decision regions for a decision rule $\hat{H} : \mathcal{Y} \to \{0, 1, 2\}$. Find $b_0(y)$, $b_1(y)$ and $b_2(y)$ — expressed in terms of $p_{Y|H}(y|i)$ and $p_H(i)$ — so that the expected penalty of the rule is given by

$$\mathbb{E}[\text{pen}(H, \hat{H}(Y))] = \sum_{y \in D_0} b_0(y) + \sum_{y \in D_1} b_1(y) + \sum_{y \in D_2} b_2(y).$$

Solution: The decision region D_j, $j = 0, 1, 2$ is given by the set of points $y \in \mathcal{Y}$ such that $\hat{H}(y) = j$. Thus we can write $\mathbb{E}[\text{pen}(H, \hat{H}(Y))]$

$$\begin{aligned} &= \sum_{y \in \mathcal{Y}} \sum_{i \in \{0,1,2\}} p_H(i) p_{Y|H}(y|i) \,\text{pen}(i, \hat{H}(y)) \\ &= \sum_{j \in \{0,1,2\}} \sum_{y \in D_j} \sum_{i \in \{0,1,2\}} p_H(i) p_{Y|H}(y|i) \,\text{pen}(i, j) \qquad [\because D_j \text{ form a partition of } \mathcal{Y}] \\ &= \sum_{j \in \{0,1,2\}} \sum_{y \in D_j} \sum_{i \in \{0,1,2\}} p_H(i) p_{Y|H}(y|i) |i - j|, \end{aligned}$$

hence we have the required relation, with $\boxed{b_j(y) = \sum_{i \in \{0,1,2\}} p_H(i) p_{Y|H}(y|i) |i - j|}$, or explicitly, $b_0(y) = p_{Y|H}(y|1) p_H(1) + 2 p_{Y|H}(y|2) p_H(2)$, $b_1(y) = p_{Y|H}(y|0) p_H(0) + p_{Y|H}(y|2) p_H(2)$, and $b_2(y) = 2 p_{Y|H}(y|0) p_H(0) + p_{Y|H}(y|1) p_H(1)$.

(b) (2 pts) Show that for any decision rule, $\mathbb{E}[\text{pen}(H, \hat{H})] \geq \sum_{y \in \mathcal{Y}} b(y)$, where

$$\begin{aligned} b(y) = \min\{ & p_{Y|H}(y|1) p_H(1) + 2 p_{Y|H}(y|2) p_H(2), \\ & p_{Y|H}(y|0) p_H(0) + p_{Y|H}(y|2) p_H(2), \\ & 2 p_{Y|H}(y|0) p_H(0) + p_{Y|H}(y|1) p_H(1) \}. \end{aligned}$$

Solution: From part (a), we have that $\mathbb{E}[\text{pen}(H, \hat{H}(Y))] = \sum_{y \in D_0} b_0(y) + \sum_{y \in D_1} b_1(y) + \sum_{y \in D_2} b_2(y)$. On each of the disjoint sets D_j, $j = 0, 1, 2$, we can lower bound the summand by $\min\{b_0(y), b_1(y), b_2(y)\}$, which is exactly $b(y)$ as stated in the question. Hence we have that

$$\begin{aligned} \mathbb{E}[\text{pen}(H, \hat{H}(Y))] &= \sum_{y \in D_0} b_0(y) + \sum_{y \in D_1} b_1(y) + \sum_{y \in D_2} b_2(y) \\ &\geq \sum_{y \in D_0} b(y) + \sum_{y \in D_1} b(y) + \sum_{y \in D_2} b(y) \\ &= \sum_{y \in \mathcal{Y}} b(y). \end{aligned}$$

(c) (2 pts) What is the decision rule that minimizes the expected penalty (in terms of $b(y)$ and/or $b_i(y)$)?

Solution: From (b), the expected penalty can never be lesser than $\sum_{y\in\mathcal{Y}} b(y)$, and this minimum is attained by choosing $\hat{H}(y) = i$ for i such that $b(y) = b_i(y)$, i.e., $\boxed{\hat{H}(y) = \arg\min_{i\in\{0,1,2\}} b_i(y)}$.

(d) (3 pts) Suppose all hypotheses are equally likely, $\mathcal{Y} = \{0, 1, 2\}$, and

$$p_{Y|H}(y|i) = \begin{cases} 0.4 & y = i, \\ 0.3 & \text{else.} \end{cases}$$

Then, what is the decision rule $\hat{H}(y)$ as an explicit function from $\mathcal{Y}$ to $\{0, 1, 2\}$ that minimizes the expected penalty? What is the MAP rule for this case?

Solution: Computing the above expression by substituting these values, we have that $\boxed{\hat{H}(y) = 1 \text{ for all } y}$ is the rule that minimizes the expected penalty. The MAP rule is given by $\boxed{\hat{H}_{\text{MAP}}(y) = \arg\max_{i\in\{0,1,2\}} p_{Y|H}(y|i) = y}$.

Remark: The MAP rule is commonly said to be "optimal", but it is important to remember in which sense — the MAP rule minimizes the *average error probability.* If we instead look to minimize some other notion of being "wrong" (such as the penalty defined in this problem), MAP is no longer optimal.

PROBLEM 2. Let $w_i(t)$, $i = 1, \ldots, m$ be the waveforms of a communication system designed for an AWGN channel with noise power spectral density $N_0/2$, suppose that $w_i(t) = 0$ whenever $t \notin [0, T]$. Let $\pi(N_0)$ denote the error probability of this system (with its optimal receiver).

Consider now two new systems:

1. The first has waveforms $\tilde{w}_i(t) = \alpha w_i(t)$, $i = 1, \ldots, m$ for a (real) scalar $\alpha \neq 0$.
2. The second has waveforms $w_i'(t) = \sum_{j=0}^{r-1} w_i(t - jT)$, $i = 1, \ldots, m$; i.e., w_i repeated r times, once every T units of time. (Here r is a positive integer.)

(a) (2 pts) How can we re-use the optimal receiver for the original system to design an optimal receiver for system 1?

Hint: Think of some pre-processing of the received signal $\tilde{R}(t)$ of system 1 before giving it as input to the optimal receiver of the original system.

Solution: Simply give $\boxed{\frac{\tilde{R}(t)}{\alpha}}$ as input to the original receiver, since $\frac{\tilde{R}(t)}{\alpha} = \frac{\tilde{w}_i(t)+N(t)}{\alpha} = w_i(t) + \frac{N(t)}{\alpha} = w_i(t) + \tilde{N}(t)$, where $\tilde{N}(t)$ is white Gaussian noise of power spectral density $\frac{N_0}{2\alpha^2}$.

(b) (2 pts) Express the error probability $\tilde{\pi}(N_0)$ of system 1 (with its optimal receiver), in terms of $\pi(N_0)$.

Solution: Since $\frac{\tilde{R}(t)}{\alpha} = w_i(t) + \tilde{N}(t)$, where $\tilde{N}(t)$ is white Gaussian noise of power spectral density $\frac{N_0}{2\alpha^2}$, we have that the error probability of this system is the same as that of the old system, except that the power spectral density of the noise is $\frac{N_0}{2\alpha^2}$ instead of $\frac{N_0}{2}$, i.e., $\boxed{\pi\left(\frac{N_0}{\alpha^2}\right)}$.

(c) (3 pts) How can we re-use the optimal receiver for the original system to design an optimal receiver for system 2?

Hint: Think of some pre-processing of the received signal $R'(t)$ of system 2 before giving it as input to the optimal receiver of the original system.

Solution: This is essentially a repetition code, and the optimum decision rule is to take the arithmetic mean: simply give $\boxed{\frac{1}{r}\sum_{j=0}^{r-1} R'(t + jT)}$ as input to the original receiver. This works because of the following: $\frac{1}{r}\sum_{j=0}^{r-1} R'(t + jT)$

$$= \frac{1}{r}\sum_{j=0}^{r-1}[w_i'(t + jT) + N(t + jT)]$$

$$= \frac{1}{r}\sum_{j=0}^{r-1}\left[\sum_{l=0}^{r-1} w_i(t + jT - lT) + N(t + jT)\right].$$

The optimal receiver of the original system computes the inner product of $R(t)$ with $w_i(t)$, which is nonzero only in $t \in [0, T]$. Hence, we only have to look at the components of $\frac{1}{r}\sum_{j=0}^{r-1} R'(t + jT)$ in $t \in [0, T]$, which is from the terms with $l = j$, given by

$$\frac{1}{r}\sum_{j=0}^{r-1}[w_i(t)+N(t+jT)] = w_i(t)+N'(t),$$

where $N'(t)$ is white Gaussian noise of power spectral density $\frac{N_0}{2r}$.

(d) (2 pts) Express the error probability $\pi'(N_0)$ of system 2 (with its optimal receiver), in terms of $\pi(N_0)$.

Solution: Since $\frac{R'(t)}{\alpha} = w_i(t) + N'(t)$, where $\tilde{N}(t)$ is white Gaussian noise of power spectral density $\frac{N_0}{2r}$, we have that the error probability of this system is the same as that of the old system, except that the power spectral density of the noise is $\frac{N_0}{2r}$ instead of $\frac{N_0}{2}$, i.e., $\boxed{\pi\left(\frac{N_0}{r}\right)}$.

Remark: As far as the error probability is concerned, repeating an input signal r times (i.e., increasing the energy by a factor of r) has the same effect as scaling the signal by $\sqrt{r}$ (leading to the energy scaling by r again).

PROBLEM 3. A bandpass transmitter for four equally likely messages is designed for an AWGN channel as follows:

- The waveform $\psi(t) = \mathrm{sinc}(t)$ is chosen as the Nyquist pulse, upon observing that $\psi(t), \psi(t-1), \psi(t-2), \dots$ form an orthonormal collection. (Note also that the Fourier transform of ψ is $\mathrm{rect}(f) = \mathbb{1}\{|f| < \frac{1}{2}\}$).
- Four codewords $c_1 = (1, 0), c_2 = (0, 1), c_3 = -c_1, c_4 = -c_2$ are chosen as vectors in $\mathbb{R}^2$. (Note: they are *real*, not complex.)
- At the transmitter, the message $i \in \{1, 2, 3, 4\}$ is first mapped to c_i, then to the baseband waveform $w_{i,E}(t) = \sum_{j=1}^{2} c_{ij}\psi(t-j)$, and finally to the transmitted waveform as

$$\begin{aligned} w_i(t) &= \sqrt{2}\Re\{w_{i,E}(t)\exp(\mathrm{j}2\pi f_c t)\} \\ &= \sqrt{2}w_{i,E}(t)\cos(2\pi f_c t) \qquad \text{with } f_c > \tfrac{1}{2}. \end{aligned}$$

At the receiver, the received signal $R(t)$ is multiplied by $\sqrt{2}\cos(2\pi f_c t)$, to form $R_E(t)$. R_E is passed through a filter with impulse response $\mathrm{sinc}(t)$, and the output of the filter is sampled at times $t_1 = 1$ and $t_2 = 2$. With Y_1 and Y_2 denoting the samples respectively, the vector $Y = (Y_1, Y_2)$ is formed. The i for which c_i is closest to Y (in the Euclidean norm) is the receiver's guess of the transmitted message.

(a) (3 pts) Is the receiver described in the above paragraph optimal? (Note: the procedure in the book would have formed the *complex* waveform $R(t)\sqrt{2}\exp(-\mathrm{j}2\pi f_c t)$ instead of the above R_E. If you claim optimality, you should explain why R_E above leads to the same decision.)

Solution: It is $\boxed{\text{optimal}}$. The book version of $R_E(t)$ is $R(t)\sqrt{2}\exp(-\mathrm{j}2\pi f_c t) = \sqrt{2}R(t)\cos(2\pi f_c t) + \mathrm{j}\sqrt{2}R(t)\sin(2\pi f_c t)$, and here, $R(t) = \sqrt{2}w_{i,E}(t)\cos(2\pi f_c t) + N(t)$, with $w_{i,E}(t)$ real. Hence, the imaginary part of the book version of $R_E(t) = \sqrt{2}R(t)\sin(2\pi f_c t) = (\sqrt{2}w_{i,E}(t)\cos(2\pi f_c t) + N(t))\sqrt{2}\sin(2\pi f_c t) = w_{i,E}(t)\sin(4\pi f_c t) + \sqrt{2}N(t)\sin(2\pi f_c t)$. When passed through the low pass filter, the first term vanishes, and we are left with just noise. Hence, the real part of the book version of $R_E(t)$ is a sufficient statistic, and this real part is exactly the $R_E(t)$ in this problem — this implies that this choice of $R_E(t)$ still leads to an optimal decision.

(b) (2 pts) What is the probability of error (in terms of N_0)?

Solution: This is simply a 4-QAM constellation through an AWGN channel, which has an error probability $\boxed{2Q\left(\frac{d}{\sqrt{2N_0}}\right) - Q\left(\frac{d}{\sqrt{2N_0}}\right)^2}$ with the minimum distance between constellation points $\boxed{d = \sqrt{2}}$.

Due to an inaccuracy in circuit design, the frequency of the cosine at the receiver is not f_c but f_c' instead, i.e., $R(t)$ is multiplied by $\sqrt{2}\cos(2\pi f_c' t)$ to form $R_E(t)$. The rest of the receiver is unchanged.

(c) (3 pts) For $x(t) = \mathrm{sinc}(t)\cos(2\pi f_0 t)$ and $y(t) = \mathrm{sinc}(t - t_0)$, show that their inner product satisfies

$$\langle x, y\rangle = \begin{cases} 0 & |f_0| \geq 1, \\ \frac{1}{2}[\mathrm{sinc}(t_0) + (1 - 2|f_0|)\,\mathrm{sinc}((2|f_0| - 1)t_0)] & \text{else.} \end{cases}$$

Hint: Use Parseval's relationship.

Solution: By Parseval's relationship, we have that $\langle x, y\rangle = \langle x_{\mathcal{F}}, y_{\mathcal{F}}\rangle$, where $x_{\mathcal{F}}(f) = \frac{\mathrm{rect}(f-f_0)+\mathrm{rect}(f+f_0)}{2}$ and $y_{\mathcal{F}}(f) = \mathrm{rect}(f)\exp(-\mathrm{j}2\pi t_0 f)$, and hence,

$$\begin{aligned}\langle x, y\rangle &= \langle x_{\mathcal{F}}, y_{\mathcal{F}}\rangle \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{rect}(f-f_0)+\mathrm{rect}(f+f_0)}{2}\,\mathrm{rect}(f)\exp(-\mathrm{j}2\pi t_0 f)\;\mathrm{d}f \\ &= \int_{-1/2}^{1/2} \frac{\mathrm{rect}(f-f_0)+\mathrm{rect}(f+f_0)}{2}\exp(-\mathrm{j}2\pi t_0 f)\;\mathrm{d}f.\end{aligned}$$

If $|f_0| \geq 1$, clearly the above integral is zero. If $|f_0| < 1$, we can continue as

$$\begin{aligned}&= \frac{1}{2}\int_{-1/2+|f_0|}^{1/2} \exp(-\mathrm{j}2\pi t_0 f)\;\mathrm{d}f + \frac{1}{2}\int_{-1/2}^{1/2-|f_0|} \exp(-\mathrm{j}2\pi t_0 f)\;\mathrm{d}f \\ &= \frac{1}{2}\int_{-1/2+|f_0|}^{1/2} \cos(2\pi t_0 f) - \mathrm{j}\sin(2\pi t_0 f)\;\mathrm{d}f \\ &\qquad + \frac{1}{2}\int_{-1/2}^{1/2-|f_0|} \cos(2\pi t_0 f) - \mathrm{j}\sin(2\pi t_0 f)\;\mathrm{d}f \\ &= \frac{1}{2}\left[\frac{\sin(2\pi f t_0)}{2\pi t_0} + \mathrm{j}\frac{\cos(2\pi f t_0)}{2\pi t_0}\right]_{-1/2+|f_0|}^{1/2} + \frac{1}{2}\left[\frac{\sin(2\pi f t_0)}{2\pi t_0} + \mathrm{j}\frac{\cos(2\pi f t_0)}{2\pi t_0}\right]_{-1/2}^{1/2-|f_0|} \\ &= \frac{1}{2}[\mathrm{sinc}(t_0) + (1-2|f_0|)\,\mathrm{sinc}((2|f_0|-1)t_0)],\end{aligned}$$

and we are done.

(d) (2 pts) Suppose $|f_c' - f_c| > 1$. What is the error probability?
Hint: Use (c) to show that Y is independent of the transmitted message.

Solution: With the mismatched frequency at the receiver, we have that $R_E(t)$ is

$$\begin{aligned}&= \sqrt{2}R(t)\cos(2\pi f_c' t) \\ &= 2w_{i,E}(t)\cos(2\pi f_c t)\cos(2\pi f_c' t) + \sqrt{2}N(t)\cos(2\pi f_c' t) \\ &= w_{i,E}(t)\cos(2\pi(f_c - f_c')t) + w_{i,E}(t)\cos(2\pi(f_c + f_c')t) + \sqrt{2}N(t)\cos(2\pi f_c' t).\end{aligned}$$

Since the matched filter is a low pass filter, the term of frequency $f_c + f_c'$ vanishes. In addition, the only terms that remain are those with frequency in $(-1/2, 1/2)$. Since $w_{i,E}$ lies in $(-1/2, 1/2)$, its product with $\cos(2\pi(f_c - f_c')t)$ is centered at $\pm|f_c' - f_c|$ and has a width of 1. Hence, if $|f_c' - f_c| > 1$, $R_E(t)$ will have no footprint in $(-1/2, 1/2)$, and hence, Y is independent of the transmitted message. The error probability is then that of a random guess between 4 choices, i.e., $\boxed{3/4}$.

(e) (2 pts) Suppose $f_c' = f_c + \frac{1}{2}$. What is the error probability?

Solution: From the calculation in (d), we have that $R_E(t) = w_{i,E}(t)\cos(\pi t) + w_{i,E}(t)\cos(2\pi(f_c + f_c')t) + \sqrt{2}N(t)\cos(2\pi f_c' t)$. Since the matched filter is a low pass filter, again, the term of frequency $f_c + f_c'$ vanishes. The output of the matched filter is then $\langle w_{i,E}(t)\cos(\pi t) + \sqrt{2}N(t)\cos(2\pi f_c' t), \mathrm{sinc}(t-j)\rangle = Y_j = d_{ij} + Z_j$, $j = 1, 2$.

The d_{ij} can be computed as

$$\begin{aligned}
\langle w_{i,E}(t)\cos(\pi t), \operatorname{sinc}(t-j)\rangle &= \left\langle \sum_{l=1}^{2} c_{il}\psi(t-l)\cos(\pi t), \operatorname{sinc}(t-j) \right\rangle \\
&= \sum_{l=1}^{2} c_{il}\langle \psi(t)\cos(\pi(t+l)), \operatorname{sinc}(t-j+l)\rangle \\
&= \sum_{l=1}^{2} c_{il}(-1)^l \langle \operatorname{sinc}(t)\cos(\pi t), \operatorname{sinc}(t-j+l)\rangle \\
&= \sum_{l=1}^{2} c_{il}(-1)^l \frac{1}{2}\operatorname{sinc}(j-l) = \frac{(-1)^j}{2} c_{ij},
\end{aligned}$$

which is still equivalent to a 4-QAM, but with distance halved. Z_j are Gaussian random variables with $\operatorname{cov}(Z_k, Z_l)$

$$\begin{aligned}
&= \frac{N_0}{2} 2\langle \cos(2\pi f_c' t)\operatorname{sinc}(t-k), \cos(2\pi f_c' t)\operatorname{sinc}(t-l)\rangle \\
&= \frac{N_0}{2}\int_{-\infty}^{\infty} 2\cos^2(2\pi f_c' t)\operatorname{sinc}(t-k)\operatorname{sinc}(t-l)\ \mathrm{d}t \\
&= \frac{N_0}{2}\int_{-\infty}^{\infty} (1+\cos(4\pi f_c' t))\operatorname{sinc}(t-k)\operatorname{sinc}(t-l)\ \mathrm{d}t \\
&= \frac{N_0}{2}\delta_{kl} + \frac{N_0}{2}\int_{-\infty}^{\infty} \cos(4\pi f_c' t)\operatorname{sinc}(t-k)\operatorname{sinc}(t-l)\ \mathrm{d}t \\
&= \frac{N_0}{2}\delta_{kl} + \frac{N_0}{2}\left\langle \cos(4\pi f_c' t)\operatorname{sinc}(t-k), \operatorname{sinc}(t-l)\right\rangle,
\end{aligned}$$

but the second term is zero, since it is the inner product of a term with frequency components centered at $|2f_c'| > 2$, and hence has no overlap with that of $\operatorname{sinc}(t-l)$. Hence the Z_j are independent Gaussian random variables with variance $\frac{N_0}{2}$. Hence the error probability of this case is the same as in (b), $\boxed{2Q\left(\frac{d}{\sqrt{2N_0}}\right) - Q\left(\frac{d}{\sqrt{2N_0}}\right)^2}$, except with the distance of constellation points halved, i.e., $\boxed{d = \frac{1}{\sqrt{2}}}$.

Remark: Modulating and demodulating correctly relies heavily on the carrier frequencies at the transmitter and receiver being matched. If the mismatch in frequency is more than 1 (as in part (d); 1 is the bandwidth of the waveform chosen), we can do no better than a random guess, which is concerning. However, in part (e), we see that even with a reasonably large mismatch of $\frac{1}{2}$, all is not lost — we get the same error probability as if the signal had been attenuated by a factor of 2, which is not terrible.

PROBLEM 4. Consider a 3-state encoding device (with states $0, 1, 2$) that accepts a sequence of data bits $b_1, b_2, \ldots$, with $b_i \in \{+1, -1\}$, and produces encoded bits $x_1, x_2, \ldots$ as follows:

Current state	Input bit b_i	Next state	Output x_{2i-1}, x_{2i}
0	$+1$	0	$+1, +1$
0	-1	1	$-1, -1$
1	$+1$	1	$-1, +1$
1	-1	2	$+1, -1$
2	$+1$	1	$+1, -1$
2	-1	0	$-1, +1$

The machine initially starts at state 0.

(a) (2 pts) After encoding k data bits $b_1, \ldots, b_k$, we would like to ensure that the machine returns to the initial state 0 by appending L termination bits $b_{k+1}, \ldots b_{k+L}$ to the data sequence. What is the value of L needed to ensure this? (Note that $(b_{k+1}, \ldots, b_{k+L})$ can depend on $(b_1, \ldots, b_k)$, but L can not.)

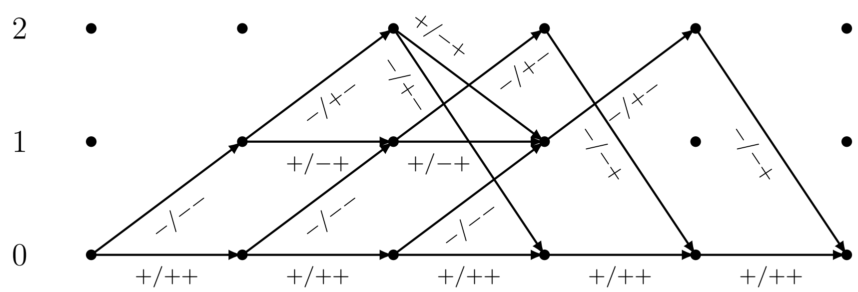
Solution: Let s_k denote the state after reading b_k. The state $s_k = 1$ requires a sequence of 2 termination bits $-, -$ to drive the state to 0. Thus we can choose $\boxed{L = 2}$, and set the termination bits to $(+, +)$, $(-, -)$, or $(-, +)$ depending on $s_k = 0, 1,$ or 2.

This encoding device, with the termination scheme in (a), is used as a transmitter for an AWGN channel. The channel output $y_1, y_2, \ldots$ is given by $y_i = \sqrt{\mathcal{E}_s} x_i + Z_i$, where $Z_1, Z_2, \ldots$ are i.i.d. $\mathcal{N}(0, \sigma^2)$.

(b) (2 pts) Draw a trellis diagram and explain how the receiver can implement the ML rule to produce $\hat{b}_1, \ldots, \hat{b}_k$ from $y_1, \ldots, y_n$.

Solution: To transmit k bits $b_1, \ldots, b_k$, the receiver has to append the $L = 2$ termination bits and send a total of $k + L$ bits, each of which produces two encoded bits. Hence the total number of y_i's needed to recover k bits is $k + L$. Having obtained $y_1, \ldots, y_n$, with $n = 2(k + L)$, the receiver simply finds the path which maximizes $\langle x, y \rangle = \sum_{i=1}^{2(k+L)} x_i y_i$. This corresponds to a branch metric of $x_{2i-1} y_{2i-1} + x_{2i} y_{2i}$ for each branch corresponding to bit i, $1 \le i \le k + L$.

The trellis diagram is shown below for $k = 3$.

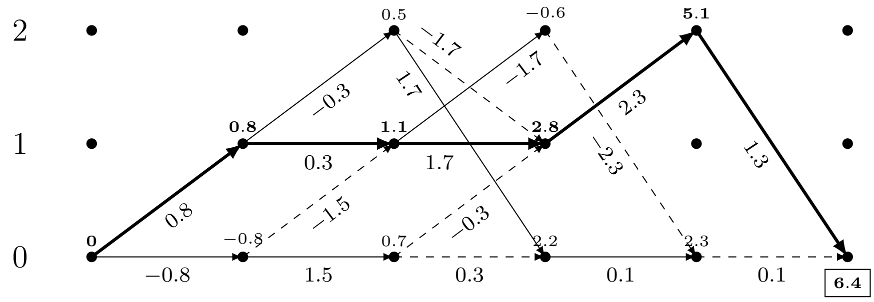


(c) (3 pts) For $k = 3$, and $y_1, y_2, \ldots$ given by

$$+0.3, -1.1, +0.6, +0.9, -0.7, +1.0, +1.2, -1.1, -0.6, +0.7, +0.4, -1.2, +0.5, -0.5, \ldots$$

determine the maximally likely $(\hat{b}_1, \hat{b}_2, \hat{b}_3)$. (You may not need the last few y_i's to do this.)

Solution: Since $k = 3$ and $L = 2$, we send a total of $k + L = 5$ bits, and hence, we need $y_1, \ldots, y_{10}$. The trellis diagram with the branch metrics evaluated is shown below. The non-surviving paths are shown with dashed lines, and the maximizing path is found by tracing back the surviving paths from the last bit.



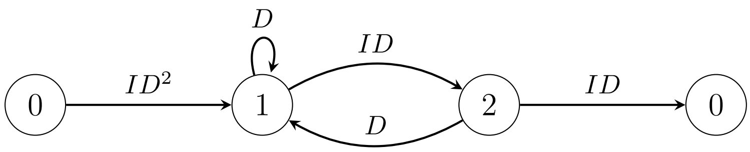
Hence, the bit sequence that is maximally likely is $\boxed{(\hat{b}_1, \hat{b}_2, \hat{b}_3) = (-1, +1, +1)}$, with $(-1, -1)$ appended to terminate at state 0.

(d) (3 pts) For the all $+1$ reference path sketch the detour flow graph labeled with $I^i D^d$'s. Construct a system of equations of the form

$$\begin{aligned} A_1(I, D) &= ?? + ??\, A_1(I, D) + ??\, A_2(I, D) \\ A_2(I, D) &= ?? + ??\, A_1(I, D) + ??\, A_2(I, D) \\ A(I, D) &= ??\, A_2(I, D), \end{aligned}$$

where A_1 and A_2 denote the transfer functions until states 1 and 2, and verify that $A(I, D) = I^3 D^4/(1 - D - ID^2)$.

Solution: The detour flow graph is shown below.



We have $A = (ID)A_2$, $A_2 = (ID)A_1$, $A_1 = (ID^2) + (D)A_1 + (D)A_2$. This yields $A(I, D) = I^3 D^4/(1 - D - ID^2)$.

(e) (2 pts) Differentiate $A(I, D)$ with respect to I, and use it to find an upper bound to the bit error probability on sending the all $+1$ sequence, as a function of $\mathcal{E}_s/\sigma^2$.

Solution: The error probability is upper bounded by, with $z = e^{-\frac{\mathcal{E}_s}{2\sigma^2}}$

$$\begin{aligned} \left.\frac{\partial A(I, D)}{\partial I}\right|_{I=1, D=z} &= \left.\frac{3I^2 D^4 (1 - D - ID^2) + I^3 D^6}{(1 - D - ID^2)^2}\right|_{I=1, D=z} \\ &= \left.\frac{3I^2 D^4 - 3I^2 D^5 - 2I^3 D^6}{(1 - D - ID^2)^2}\right|_{I=1, D=z} \\ &= \boxed{\frac{z^4 (3 - 3z - 2z^2)}{(1 - z - z^2)^2}}. \end{aligned}$$

Remark: This is not a convolutional encoder as seen in the course, but it is still a finite-state encoder. Thus, we can still draw trellis diagrams and use them to efficiently carry out ML decoding. One difference, however, is that it does not have the symmetry of a convolutional code. The upper bound computed in part (e) is only an upper bound to the error probability on sending the all $+1$ sequence, and not the average error probability (as is the case for convolutional codes). An interesting exercise would be to repeat parts (d) and (e) taking some other path as the reference — the transfer function (and hence the upper bound derived from it) will depend on the reference path (unlike convolutional codes, which have more structure).