EE121 Midterm Solution

by Lizhong Zheng

Problem 1 (a) We can use the binary tree to construce a mapping from code words to subintervals of $[0,1)$ as following: for a code word $b_1 b_2 \dots b_l$, let $s = (0.b_1 b_2 \dots b_l)_b$ be the number in $[0,1)$ whose binary expansion is the code word. Assign to the code word the subinterval $[s, s+2^{-l})$. For example,

$$\begin{aligned} '0' &\rightarrow [0, \frac{1}{2}) \\ '101' &\rightarrow [\frac{5}{8}, \frac{6}{8}) \end{aligned}$$

Now we claim that for any prefix free code, all the subintervals are non-overlapping. To see this, if two subintervals I_1, I_2 are overlapping, this means that at least one of the start points lies in the other subinterval. Assume the CW of I_1 is $b_1 b_2 \dots b_{l_1}$, and the CW of I_2 is $c_1 c_2 \dots c_{l_2}$. w.o.l.g. assume that

$$(0.c_1 c_2 \dots c_{l_2})_b \in I_1$$

but this means that $b_1 b_2 \dots b_{l_1}$ is a prefix of $c_1 c_2 \dots c_{l_2}$. Thus we proved the claim.

Now from this result, we know the sub-intervals of length 2^{-l_i} are non-overlapping, and the total length $\sum_i 2^{-l_i} \leq 1$.

(b) True. Since for optimal binary code, the corresponding binary tree can not have any empty leaf. This is because in binary tree, one leaf can have only one brother. Thus if there is any empty leaf, we can always move it's brother branch up to their parent node, which will reduce the code word length. It is obvious by the mapping in part (a) that for a tree with no empty leaf, the corresponding code has $\sum_i 2^{-l_i} = 1$.

(c) For $D-ary$ Huffman code, the Kraft's inequality may not hold with equality. Since when $D > 2$, each leaf of the tree has more than 1 brothers. Thus one empty leaf may not be sufficient to move it's brother up as we did in part (b). An easy counter example is if $D = 3$, and we only have 2 symbols to code.

(d) Let the alphabet size be N, we order the letters with probability $2^{-k_i}, i = 1, \dots N$ to have

$$k_1 \leq k_2 \leq \dots \leq k_N$$

For each i, assign to letter i the code word as the binary expansion of $\sum_{j=1}^{i-1} 2^{-k_j}$ with length k_i. Since the sub-intervals corresponding to each CW has length 2^{-k_i}, and the subintervals are non-overlapping. This code is prefix free. In fact, this is a Huffman code.

The optimality of this code comes from the fact that $E[L] = \sum k_i 2^{-k_i} = H(X)$.

Problem 2

(a) The p.s.d $S_X(f) = \mathcal{F}[R_X(\tau)] = \frac{2}{1+4\pi^2 f^2}$.

(b) Let the sampled process by $Y_n = X(nT)$.

$\{Y_n\}$ is a Gaussian process, since any subset of it is a set of samples of $X(t)$, which by definition is jointly Gaussian.

$\{Y_n\}$ is WSS. To see this

$$\begin{aligned} R_Y(n, n+k) &= E[Y_n Y_{n+k}] \\ &= E[X(nT)X(nT+kT)] = R_X(kT) \end{aligned}$$

only depends on k. Thus the process $\{Y_n\}$ is WSS, and the autocorrelation function is $R_Y(k) = e^{-|k|T}$.

(c) The LLSE

$$\begin{aligned} \hat{Y}_n|Y_{n-1} &= E[Y_n Y_{n-1}]E[Y_{n-1}^2]^{-1}Y_{n-1} \\ &= e^{-T}Y_{n-1} \end{aligned}$$

The estimation error $Y_n - \hat{Y}_n$ is Gaussian distributed, since it is a linear combination of Y_n and Y_{n-1}, which are joint Gaussian. Therefore, the distribution is decided by the mean and variance.

$$\begin{aligned} E[Y_n - \hat{Y}_n] &= E[Y_n - e^{-T}Y_{n-1}] = 0 \\ E[(Y_n - \hat{Y}_n)^2] &= E[(Y_n - e^{-T}Y_{n-1})^2] \\ &= R_Y(0) + e^{-2T}R_Y(0) - 2e^{-T}R_Y(1) \\ &= 1 - e^{-2T} \end{aligned}$$

Therefore, we have $Y_n - \hat{Y}_n \sim N(0, 1 - e^{-2T})$.

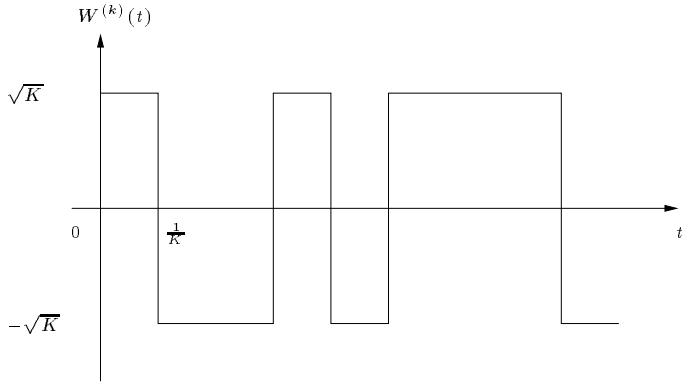
(d) Write $Z_n = Y_n - \hat{Y}_n$. To minimize the quantization error, we take $\Delta = E[Z_n \mid Z_n \geq 0]$, thus

$$\begin{aligned} \Delta &= 2\int_0^\infty t \frac{1}{\sqrt{2\pi}}\sigma e^{-\frac{t^2}{2\sigma^2}}\, dt \\ &= \frac{2\sigma}{\sqrt{2\pi}} \end{aligned}$$

where $\sigma_2 = 1 - e^{2T}$. As $T \to$, the samples Y_n are highly correlated, thus the LLSE is more and more precise, so $\Delta \to 0$.

Problem 3

(a)



(b)

$$\begin{aligned} Var[W^{(k)}(t)] &= Var[W_n\sqrt{K}] = KVar[W_n] = K \\ Var[\int_0^1 W^{(k)}(t)dt] &= Var[\sum_{i=1}^{K} \frac{1}{K}\sqrt{K}W_i] \\ &= 1 \end{aligned}$$

Thus, as $K \to \infty$, the total power of process $W^{(k)}$ approaches infinity, but the power along the direction of function $\phi(t) = U(t) - U(t-1)$ is finite. The integral $\int_0^1 W^{(k)}(t)dt$ can be written as averaging K iid random variables, $W_1, \ldots, W_K$. If you have learnt Central Limit Theorem, you will see that the scaling factor $\sqrt{K}$ is the only right scaling to keep the power of the filtered process to be 1. In fact, one can show that the power of $W^{(k)}$ projected to any direction is 1, ie. if $\int_0^1 |\phi(t)|^2 dt = 1$,

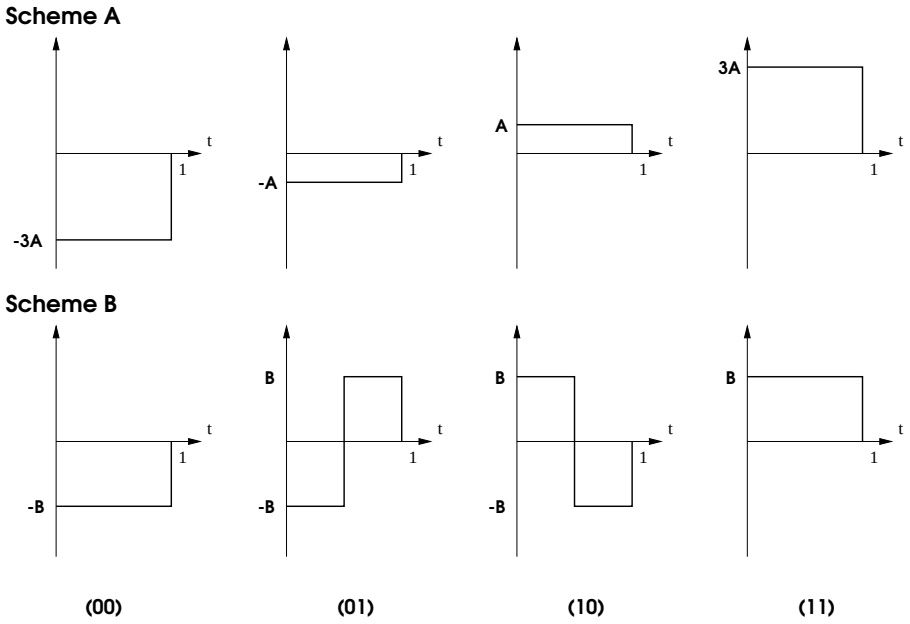
$$Var[\int_0^1 W^{(k)}(t)\phi(t)\ dt] = 1$$

Thus as K increases, the power per dimension is fixed, but the process occupies more and more dimensions, thus the total power increases with no bound.

(c) As $K \to 0$, the resulting process occupies ∞ dimensions, thus if we fix the total power of the process, then projected to any finite dimensional subpace, the process has 0 power. This means if we use linear filter to reduce the noise to the signal space that we are interested in, then the filtered noise is always 0, which is not very useful.

Problem 4

(a) The waveforms are shown as following



(b) For Scheme A, the signal space has dimension 1, the basis function is $\phi(t) = 1$ for $t \in [0, 1)$.

For Scheme B, the signal space has dimension 2, one basis is

$$\begin{aligned} \phi_1(t) &= \sqrt{2} \text{ for } t \in [0, \frac{T}{2}) \\ &\quad 0 \text{ otherwise} \\ \phi_2(t) &= \sqrt{2} \text{ for } t \in [\frac{T}{2}, T) \\ &\quad 0 \text{ otherwise} \end{aligned}$$

The constellation w.r.t the basis above is shown in the figure in the last page.

(c) For scheme A, the minimum distance is $2A = d$, thus $A = \frac{d}{2}$, the average power transmitted is

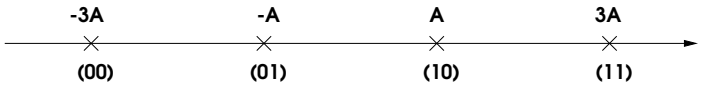
$$\begin{aligned} P_A &= 2 \times \frac{1}{4}A^2 + 2 \times \frac{1}{4}(3A^2) \\ &= \frac{5}{4}d^2 \end{aligned}$$

For scheme B, the minimum distance is $2B/\sqrt{2} = d$, thus $B = \frac{d}{\sqrt{2}}$, the average power transmitted is

$$P_B = B^2 = \frac{1}{2}d^2$$

Thus Scheme B is better, since it requires less power to achieve the same minimum distance.

Scheme A



Scheme B