1 Neural Network-Based Soft-Output Detector

Using the hyperparameters given in Table 1 with the developed neural network (see the zip file for solutions) framework from Tutorial 2

Table 1: Parameters for Tweaking

Parameter	Value
num_epochs	500
beta	1.6
N_training	2^{12}
hidden_size1	10
hidden_size2	10
retrain_network	True

we get the following Average Information Rate (AIR) for different values of SNR on a logarithmic scale

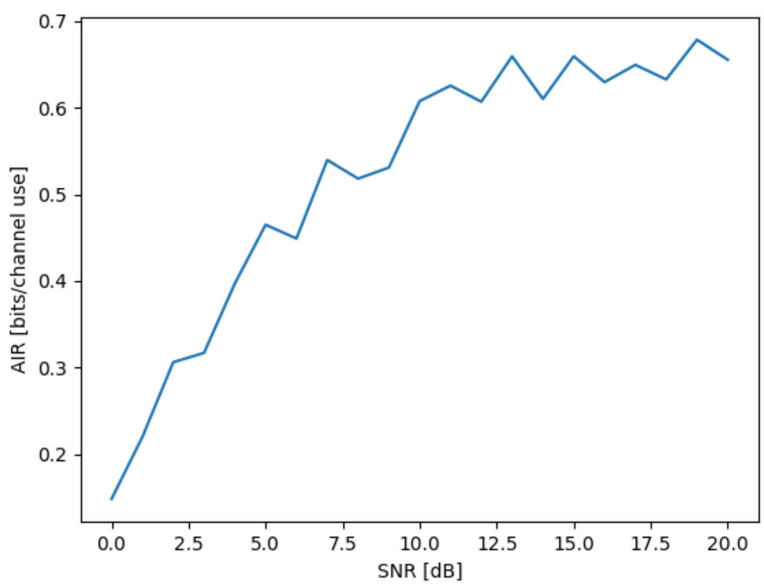

Figure 1: AIR vs SNR using NN-based soft-detector

Since the purpose is soft-decoding, it needs to calculate the probability of each codeword given a sequence of observations, we used the softmax function. The Softmax function allows us to represent the output of the NN as a probability for every class (codeword). The Softmax function does that by first mapping it to a positive value by taking exponential and normalizing the other classes. Therefore, deploying softmax generates a probability mass function at the output. As it can be seen from Figure 1, AIR is very close to 1, which is the maximum rate for BPSK, also it saturates for high SNR values reaching a value of 0.7 approximately.

2 Sum-Product Based Soft-Output Detector

a)

$$\underline{\hat{x}} = \underset{\underline{x}}{\text{argmax}}\ p(\underline{x}, \underline{y}, \underline{h}) \tag{1}$$

b)

$$\underline{\hat{x}} = \underset{\underline{x}}{\text{argmax}} \prod_{i=1}^{k} P(x_i) \cdot p(h_i \mid h_{i-1})\, p(y_i \mid x_i, h_i) \tag{2}$$

The derivation of (2) follows from the Bayes Rule and conditional independence and also channel is the Gaussian-Markov process. Namely,

$$\begin{aligned}
\underline{\hat{x}} &= \underset{\underline{x}}{\text{argmax}}\ p(\underline{x}, \underline{y}, \underline{h}) \\
\underline{\hat{x}} &= \underset{\underline{x}}{\text{argmax}}\ p(\underline{x}, \underline{h}) \cdot p(\underline{y} \mid \underline{x}, \underline{h}) \\
\underline{\hat{x}} &= \underset{\underline{x}}{\text{argmax}}\ P(\underline{x}) \cdot p(\underline{h}) \cdot p(\underline{y} \mid \underline{x}, \underline{h}) \\
\underline{\hat{x}} &= \underset{\underline{x}}{\text{argmax}} \prod_{i=1}^{k} P(x_i) \cdot p(\underline{h}) \cdot p(\underline{y} \mid \underline{x}, \underline{h}) \\
\underline{\hat{x}} &= \underset{\underline{x}}{\text{argmax}} \prod_{i=1}^{k} P(x_i) \cdot p(h_i \mid h_{i-1})\, p(y_i \mid x_i, h_i)
\end{aligned}$$

by using the knowledge that $\underline{x}$ and $\underline{h}$ are independent, entries of $\underline{x}$ are independent, the channel $\underline{h}$ is a Gaussian-Markov process, therefore, it only depends on the previous channel entry, so we can write it as multiplication of conditional densities, knowing $\underline{x}$ and $\underline{h}$, then y_i values are also independent. We obtain the final form (2).

c)

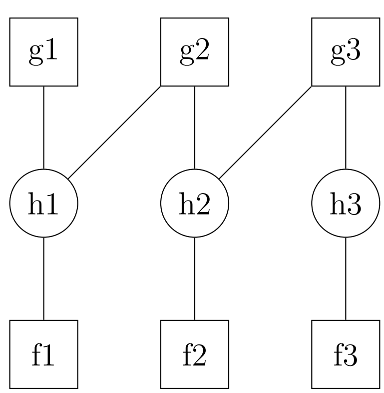

Figure 2: Factor graph of the joint density

The factor graph in Figure 2 is derived by using the given notation of

$$\begin{aligned} f_i\left(h_i\right) &= P\left(x_i\right) p\left(y_i \mid x_i, h_i\right) \\ g_i\left(h_i h_{i-1}\right) &= p\left(h_i \mid h_{i-1}\right). \end{aligned}$$

When we use the given notation, we reformulate the maximization objective in (2) as

$$f_1\left(h_1\right) \cdot g_1\left(h_1\right) \cdot f_2\left(h_2\right) \cdot g_2\left(h_2 h_1\right) \cdot f_3\left(h_3\right) \cdot g_3\left(h_3 h_2\right), \tag{3}$$

and whose factor graph is given in Figure 2.

d)

$$m_{g_1 \to h_1}\left(h_1\right) = p\left(h_1\right) \tag{4}$$

$$m_{g_2 \to h_2}\left(h_2\right) = p\left(x_1, y_1, h_2\right) \tag{5}$$

$$m_{g_3 \to h_3}\left(h_3\right) = p\left(x_1, y_1, x_2, y_2, h_3\right) \tag{6}$$

Derivation of (5) is given as

$$\begin{aligned} m_{g_2 \to h_2}\left(h_2\right) &= \sum_{h_1} g_2\left(h_2 h_1\right) m_{h_1 \to g_2}\left(h_1\right) \\ &= \sum_{h_1} g_2\left(h_2 h_1\right) m_{l_1 \to h_1}\left(h_1\right) m_{g_1 \to h_1}\left(h_1\right) \\ &= \sum_{h_1} g_2\left(h_2 h_1\right) f_1\left(h_1\right) g_1\left(h_1\right) \\ &= \sum_{h_1} p\left(h_2 \mid h_1\right) P\left(x_1\right) p\left(y_1 \mid x_1, h_1\right) p\left(h_1\right) \\ &= \sum_{h_1} p\left(h_2 \mid h_1\right) p\left(x_1, y_1 \mid h_1\right) p\left(h_1\right) \\ &= \sum_{h_1} p\left(x_1, y_1, h_2 \mid h_1\right) p\left(h_1\right) \\ &= \sum_{h_1} p\left(x_1, y_1, h_1, h_2\right) \\ &= p\left(x_1, y_1, h_2\right) \end{aligned}$$

Derivation of (6) is given as

$$\begin{aligned}
m_{g_3 \to h_3}(h_3) &= \sum_{h_2} g_3(h_3, h_2)\, m_{h_2 \to g_3}(h_2) \\
&= \sum_{h_2} g_3(h_3 h_2)\, m_{g_2 \to h_2}(h_2)\, m_{g_2 \to h_2}(h_2) \\
&= \sum_{h_2} p(h_3 \mid h_2)\, p(x_1, y_1, h_2)\, P(x_2)\, p(y_2 \mid x_2, h_2) \\
&= \sum_{h_2} p(h_3 \mid h_2)\, p(x_1, y_1, h_2)\, p(x_2, y_2 \mid h_2) \\
&= \sum_{h_2} p(x_1, y_1, x_2, y_2, h_3 \mid h_2)\, p(h_2) \\
&= p(x_1, y_1, x_2, y_2, h_3)
\end{aligned}$$

e)

$$\int_{-\infty}^{\infty} \prod_{i=1}^{3} \mathcal{N}\left(x; \mu_i, \sigma_i^2\right) \mathrm{d}x = \mathcal{N}\left(\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2\right) \mathcal{N}\left(\frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}; \mu_3, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} + \sigma_3^2\right) \tag{7}$$

The product of two Gaussian PDFs is a weighted Gaussian PDF with a weight of Gaussian function. This is given in the first page of the footnote.[1] Namely,

$$\mathcal{N}\left(x; \mu_1, \sigma_1^2\right) \cdot \mathcal{N}\left(x; \mu_2, \sigma_2^2\right) = \frac{1}{\sqrt{2\pi}\sigma_{12}} \exp\left[-\frac{(x-\mu_{12})^2}{2\sigma_{12}^2}\right] \frac{1}{\sqrt{2\pi\left(\sigma_1^2 + \sigma_2^2\right)}} \exp\left[-\frac{(\mu_1 - \mu_2)^2}{2\left(\sigma_1^2 + \sigma_2^2\right)}\right] \tag{8}$$

$$= \mathcal{N}\left(\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2\right) \cdot \mathcal{N}\left(x; \mu_{12}, \sigma_{12}^2\right) \tag{9}$$

with

$$\sigma_{12} = \sqrt{\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}} \text{ and } \mu_{12} = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

Since $\mathcal{N}\left(\mu_1; \mu_2, \sigma_1^2 + \sigma_2^2\right)$ is independent of x, we can take it out of the infinite integral in (7). Then we apply again the product of two Gaussian PDFs to reach (7). Namely,

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x; \mu_{12}, \sigma_{12}^2\right) \cdot \mathcal{N}\left(x; \mu_3, \sigma_3^2\right) \mathrm{d}x = \mathcal{N}\left(\mu_{12}; \mu_3, \sigma_{12}^2 + \sigma_3^2\right) \cdot \int_{-\infty}^{\infty} \mathcal{N}\left(x; \mu_{123}, \sigma_{123}^2\right) \mathrm{d}x. \tag{10}$$

Since the integral is over a PDF, it equals 1, only $\mathcal{N}\left(\mu_{12}; \mu_3, \sigma_{12}^2 + \sigma_3^2\right)$ remains. Therefore, (7) is proven.

[1] The derivation is given here

f)

$$\mathcal{N}\left(ax;\mu,\sigma^2\right)=\mathcal{N}\left(a\mu;x,\sigma^2\right)\quad a\in\{-1,1\} \tag{11}$$

Since for Gaussian PDF we can exchange mean and the variable due to symmetry, we can show (11) by plugging the values of a. It is trivial to show the case for $a=1$ because just because of symmetry the two terms are equal and for $a=-1$ because of the square in the exponent, the two terms are still equal.

g)

$$\begin{aligned}
&P(\underline{x},\underline{y})=\frac{1}{8}N\left(\mu_h^2;x_1y_2,\sigma_h^2+\sigma_0^2\right)N\left(\delta;x_2y_2,\varphi^2+\sigma_0^2\right)N\left(\alpha;x_3y_3,\beta^2+\sigma_0^2\right)\\
&\delta=\frac{\mu_h\sigma_0^2+x_1y_1\sigma_h^2}{\sigma_0^2+\sigma_h^2},\varphi^2=\frac{\sigma_0^2\sigma_h^2}{\sigma_0^2+\sigma_h^2}+\sigma_0^2\\
&\alpha=\frac{\delta\sigma_0^2+x_2y_2\varphi^2}{\sigma_0^2+\varphi^2},\beta^2=\frac{\sigma_0^2\varphi^2}{\sigma_0^2+\varphi^2}+\sigma_0^2
\end{aligned} \tag{12}$$

To derive it, we use the sum-product algorithm and simplify the equations by using equalities we showed in subproblems e) and f).

$$\begin{aligned}
&P\left(\underline{x},\underline{y},h_3\right)=m_{g_3\to h_3}\left(h_3\right)m_{P_3\to h_3}\left(h_3\right)\\
&P(\underline{x},\underline{y})=\int_{h_3}P\left(\underline{x},\underline{y},h_3\right)dh_3=\int_{h_3}m_{g_3\to h_3}\left(h_3\right)m_{f_3\to h_3}\left(h_3\right)dh_3\\
&P(\underline{x},\underline{y})=\int_{h_3}\int_{h_2}g_3\left(h_3h_2\right)m_{g_2\to h_2}\left(h_2\right)m_{f_2\to h_2}\left(h_2\right)m_{f_3\to h_3}\left(h_3\right)dh_2dh_3\\
&P(\underline{x},\underline{y})=\int_{h_3}\left[\int_{h_2}g_3\left(h_3h_2\right)\left[\int_{h_1}g_2\left(h_2h_1\right)m_{f_1\to h_1}\left(h_1\right)m_{g_1\to h_1}\left(h_1\right)dh_1\right]m_{f_2\to h_2}\left(h_2\right)dh_2\right]m_{f_3\to h_3}\left(h_3\right)dh_3
\end{aligned}$$

We can compute this integral since now we have a model for every term inside the integral. Namely,

$$\begin{aligned}
&m_{h_1\to h_1}\left(h_1\right)=f_1\left(h_1\right)=P\left(x_1\right)p\left(y_1\mid x_1,h_1\right)=\frac{1}{2}N\left(y_1;h_1x_1,\sigma_0^2\right)\\
&m_{g_1\to h_1}\left(h_1\right)=g_1\left(h_1\right)=p\left(h_1\right)=N\left(h_1;\mu_h;\sigma_n^2\right)\\
&m_{f_2\to h_2}\left(h_2\right)=f_2\left(h_2\right)=P\left(x_2\right)p\left(y_2\mid x_1,h_2\right)=\frac{1}{2}N\left(y_2;h_2x_2,\sigma_0^2\right)\\
&m_{g_3\to h_3}\left(h_3\right)=h_3\left(h_3\right)=P\left(x_3\right)P\left(y_3\mid x_3,h_3\right)=\frac{1}{2}N\left(y_3;h_3x_3,\sigma_0^2\right)\\
&g_3\left(h_3h_2\right)=p\left(h_3\mid h_2\right)=\frac{1}{2}N\left(h_3;h_2,\sigma_n^2\right)\\
&g_2\left(h_2h_1\right)=p\left(h_2\mid h_1\right)=\frac{1}{2}N\left(h_2;h_1,\sigma_h^2\right)
\end{aligned}$$

Plugging all the modeled densities into the joint density, and using the fact that all transmit symbols are equally likely takes values of either 1 or -1, namely, $P\left(\underline{x}\right)$, we get the following integral

$$P(\underline{x}, \underline{y}) = \frac{1}{8} \int_{h_3} \int_{h_2} N\left(h_3; h_2, \sigma_h^2\right) \int_{h_1} N\left(h_2; h_1, \sigma_h^2\right) N\left(y_1; h_1 x_1, \sigma_0^2\right) N\left(h_1; \mu_1 \sigma_n^2\right) dh_1$$
$$N\left(y_2; x_2 h_2, \sigma_0^2\right) dh_2 N\left(y_3; x_3 h_3, \sigma_0^2\right) dh_3$$

Let's start solving the integrals starting from most inside one. Namely, we solve

$$\int_{h1} N\left(h_2; h_1, \sigma_n^2\right) N\left(y_1; h_1 x_1, \sigma_0^2\right) N\left(h_1; \mu_h, \sigma_h^2\right) dh_1 = \int_{h1} N\left(h_1; h_2, \sigma_h^2\right) N\left(h_1; x_1 y_1, \sigma_0^2\right) N\left(h_1; \mu_h, \sigma_h^2\right) dh_1$$

To write this integral, we used the symmetry of Gaussian PDF, symmetry allows us to change the place of the mean and the variable. We also used the equality given in the subproblem (11) to get $N\left(y_1; h_1 x_1, \sigma_0^2\right) = N\left(h_1; x_1 y_1, \sigma_0^2\right)$. Now, we can apply the equality in (7), since the integral comprises the multiplication of three Gaussian PDFs with the same variable. For the sake of easiness, we define

$$\begin{array}{ll} \mu_1 \triangleq \mu_h & \sigma_1^2 \triangleq \sigma_h^2 \\ \mu_2 \triangleq x_1 y_1 & \sigma_2^2 \triangleq \sigma_0^2 \\ \mu_3 \triangleq h_2 & \sigma_3^2 = \sigma_n^2 \end{array} \tag{13}$$

Now, we can insert these definitions into (7) and get the following integral

$$P(\underline{x}, \underline{y}) = \frac{1}{8} N\left(\mu_h; x_1 y_1, \sigma_h^2 + \sigma_0^2\right) \int_{h_3} \int_{h_2} N\left(\frac{\mu_h \sigma_0^2 + x_1 y_1 \sigma_h^2}{\sigma_0^2 + \sigma_h^2}; h_2, \frac{\sigma_0^2 \sigma_h^2}{\sigma_0^2 + \sigma_h^2} + \sigma_h^2\right) N\left(h_2; h_3, \sigma_h^2\right)$$
$$N\left(h_2; x_2 y_2, \sigma_0^2\right)\ dh_2\ N\left(h_3; x_3 y_3, \sigma_0^2\right)\ dh_3, \tag{14}$$

where we again used the (7) and (11). Going further we can solve the inner integral in the same way, using (7). But for the sake of easiness, let us define the mean and variances again as

$$\begin{array}{ll} \mu_1 \triangleq \delta \triangleq \frac{\mu_h \sigma_0^2 + x_1 y_1 \sigma_h^2}{\sigma_0^2 + \sigma_h^2} & \sigma_1^2 \triangleq \varphi^2 \triangleq \frac{\sigma_0^2 \sigma_h^2}{\sigma_0^2 + \sigma_h^2} + \sigma_h^2 \\ \mu_2 \triangleq x_2 y_2 & \sigma_2^2 \triangleq \sigma_0^2 \\ \mu_3 \triangleq h_3 & \sigma_3^2 \triangleq \sigma_h^2 \end{array} \tag{15}$$

By using these means and variances, we can use (7) again to construct

$$\begin{aligned} &\int_{h_2} N\left(h_2; \delta, \varphi^2\right) N\left(h_2; h_3, \sigma_h^2\right) N\left(h_2; x_2 y_2, \sigma_0^2\right) dh_2 \\ &= N\left(\delta; x_2 y_2, \varphi^2 + \sigma_0^2\right) N\left(\frac{\delta \sigma_0^2 + x_2 y_2 \varphi^2}{\sigma_0^2 + \varphi^2}; h_3, \frac{\sigma_0^2 \varphi^2}{\sigma_0^2 + \varphi^2} + \sigma_n^2\right). \end{aligned} \tag{16}$$

Then, by inserting (16) into (15), we obtain

$$P(\underline{x}, \underline{y}) = \frac{1}{8} N\left(\mu_h; x_1 y_1, \sigma_h^2 + \sigma_0^2\right) N\left(\delta; x_2 y_2, \varphi^2 + \sigma_0^2\right) \int_{h_3} \mathcal{N}\left(h_3; \alpha, \beta^2\right) \mathcal{N}\left(h_3; x_3 y_3, \sigma_0^2\right) dh_3, \tag{17}$$

where we defined $\alpha \triangleq \frac{\delta\sigma_0^2+x_2y_2\varphi^2}{\sigma_0^2+\varphi^2}$ and $\beta \triangleq \frac{\sigma_0^2\varphi^2}{\sigma_0^2+\varphi^2}+\sigma_n^2$.

Finally, only one integral remains to be solved. However, we cannot directly use (7), since in (17), there are only two Gaussian PDFs instead of three. Luckily, we already derived it for two Gaussian cases also in (10). Therefore, the inside integral is equal to $\mathcal{N}\left(\alpha; x_3y_3, \beta^2+\sigma_0^2\right)$ and the final result of (12) is obtained.

h)

Looking at (12), we see that for every element in the sequence, a new Gaussian term whose mean and variance depend on the previous term is added. Following this pattern, we come up with the following algorithm Given $\underline{x}, \sigma_h^2, \sigma_0^2, \mu_h, \underline{y}$

$$p(\underline{x},\underline{y}) = \frac{1}{2^k}\prod_{i=1}^{k} N\left(\mu_{(i)}; x_iy_i, \sigma_{(i)}^2\right)$$

where $\mu_{(3)} = \mu_h, \sigma_{(1)}^2 = \sigma_h^2$

$$\begin{aligned}
&\mu_{(1)} = \mu_h, \sigma_{(1)}^2 = \sigma_h^2 \\
&\mu_{(2)} = \frac{\mu_{(1)}\sigma_0^2 + x_1y_1\sigma_{(1)}^2}{\sigma_0^2+\sigma_{(1)}^2}, \sigma_{(2)}^2 = \frac{\sigma_0^2\sigma_{(1)}^2}{\sigma_0^2+\sigma_{(1)}^2} + \sigma_0^2 + \sigma_h^2 \\
&\vdots \\
&\mu_{(i+1)} = \frac{\mu_{(i)}\sigma_0^2 + x_iy_i\sigma_{(i)}^2}{\sigma_0^2+\sigma_{(i)}^2}, \sigma_{(i+1)}^2 = \frac{\sigma_0^2\sigma_{(i)}^2}{\sigma_0^2+\sigma_{(i)}^2} + \sigma_0^2 + \sigma_h^2
\end{aligned}$$

i)

Putting $\underline{x}$ into matrix format by stacking different transmitted codewords as columns. We can run the algorithm. However, we don't want to optimize over the joint density but over the posterior distribution given as

$$p(\underline{x} \mid \underline{y}) = \frac{p(\underline{x},\underline{y})}{\sum_{\underline{x}'\in\{-1,1\}^k} p\left(\underline{x}',\underline{y}\right)}, \tag{18}$$

since it shrinks the solution space that we are dealing with.

j) and k)

As can be seen in Figure 3, AIR vs SNR performance is much more stable and much higher than NN-based soft decoder. This is because, instead of inferring the probability distribution from data in a supervised way, we use the known statistics of the individual components and calculate them effectively using the sum-product algorithm. In the best-case scenario, our NN-based soft decoder would perform the same as the sum-product algorithm, since it is the best estimator that we got using the maximum a posteriori estimates.

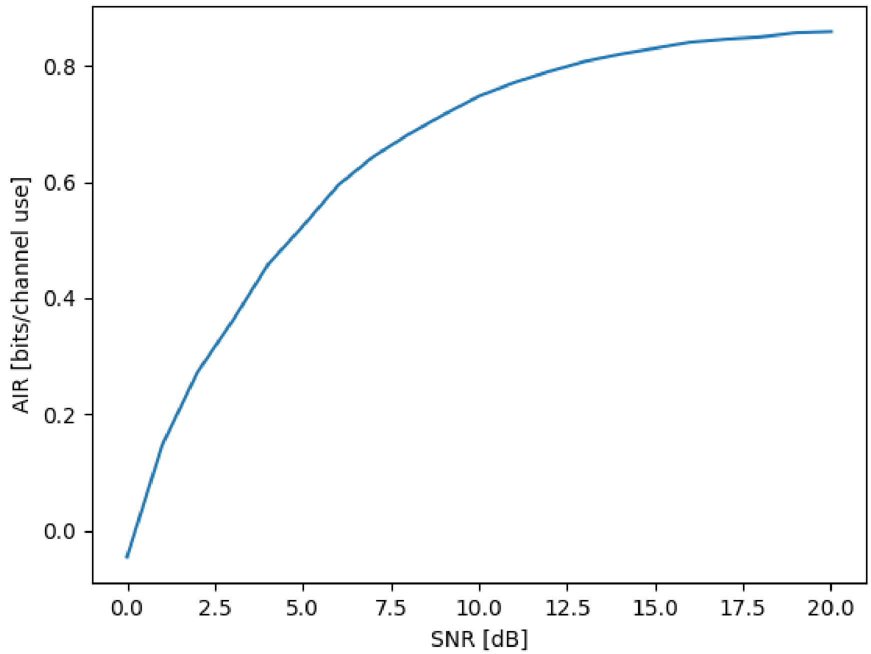

Figure 3: AIR vs SNR using sum-product based soft-detector

3 Expectation Maximization

E-step:

$$Q^{(t)}_{X_i|Y_i}(x \mid y_i) = \frac{p^{(t)}_j \frac{1}{\sigma^2_{(t)}} \exp\left(-\frac{\left|y_i - \Delta^{(t)} x\right|^2}{\sigma^2(t)}\right)}{\sum_{k \in X} P^{(t)}_k \frac{1}{\sigma^2_{(t)}} \exp\left(-\frac{\left|y_i - \Delta^{(t)} x_k\right|^2}{\sigma^2(t)}\right)} \tag{19}$$

M-step:

$$\Delta^{\mathrm{EM}}_{\mathrm{opt}} = \frac{\sum_{i=1}^{n} \sum_{x \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}(x \mid y_i) y_i x^*}{\sum_{i=1}^{N} \sum_{x \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}(x \mid y_i) |x|^2}, \tag{20}$$

$$\sigma^{2,\mathrm{EM}}_{\mathrm{opt}} = \frac{2}{n} \sum_{i=1}^{n} \sum_{x \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}(x \mid y_i) \left|y_i - \Delta^{\mathrm{EM}}_{\mathrm{opt}} x\right|^2 \tag{21}$$

$$p^{\mathrm{EM}}_{j,\mathrm{opt}} = \frac{1}{n} \sum_{i=1}^{n} Q^{(t)}_{X_i|Y_i}(x_j \mid y_i) \tag{22}$$

Let us derive the equations given in 19-22. For the E-step, we start by introducing auxiliary distributions to make KL-divergence zero. Also, in the system model, it is given that receiver noise is zero mean circularly-symmetric complex Gaussian with unknown variance σ^2. The

PDF of $y_i \mid x$ is therefore also circularly-symmetric complex Gaussian with a mean of Δx and variance σ^2. The PDF of $y_i \mid x$ is given as

$$p_{Y|X}\left(y_i \mid x; \underline{\theta}\right) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{\left|y_i - \Delta x\right|^2}{\sigma^2}\right). \tag{23}$$

Since we also don't have Δ, σ^2, and PMF of x, we parametrize this PDF with $\underline{\theta} = [\Delta, \sigma^2, p(\underline{x})]$, and get (19). Now for the M-step, we form the optimization problem

$$\underline{\theta}^{(t+1)} = \underset{\underline{\theta}}{\text{armax}} \sum_{i=1}^{n} \sum_{j \in x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \ln p_{Y_i X_i}\left(y_i, x_j; \underline{\theta}\right) \text{ s. t. } p(\underline{x}) \text{ is a PMF.} \tag{24}$$

When we apply Bayes' rule to (24) and apply (23) into (24), we get

$$\sum_{i=1}^{n} \sum_{x_j \in \mathcal{X}} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \left(-\ln\left(\pi\sigma^2\right) - \frac{\left|y_i - \Delta x_j\right|^2}{\sigma^2} + \ln p_j\right). \tag{25}$$

If we take the partial derivative of (25) with respect to the complex variable Δ using Wirtinger derivates [2], that is the partial derivative of Δ^* with respect to Δ is zero. Namely,

$$\begin{aligned}
&\sum_{i=1}^{n} \sum_{x_j + x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \frac{\partial}{\partial \Delta}\left(-\frac{\left|y_i - \Delta x_j\right|^2}{\sigma^2}\right) \stackrel{!}{=} 0 \\
&\sum_{i=1}^{n} \sum_{x_j + x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \frac{\partial}{\partial \Delta}\left(\left(y_i i_i^*\right) - y_i \Delta^* x_j^* - y_i^* \Delta x_j + \Delta\Delta^* x_j^* x_j\right) \stackrel{!}{=} 0 \\
&\sum_{i=1}^{n} \sum_{x_j \pm x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \left(x_j y_i^* - \left|x_j\right|^2 \Delta^*\right) \stackrel{!}{=} 0 \\
&\Delta^* = \frac{\sum_{i=1}^{n} \sum_{x_j + x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) x_j y_i^*}{\sum_{i=1}^{n} \sum_{x_j \in x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \left|x_j\right|^2} \Rightarrow \Delta = \frac{\sum_{i=1}^{n} \sum_{x_j \in x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) y_i x_j^*}{\sum_{i=1}^{n} \sum_{x_j \in x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \left|x_j\right|^2}
\end{aligned}$$

Then, to derive (21), we take derivative with respect to σ^2, and only contributing terms are

$$\begin{aligned}
&\frac{\partial}{\partial \sigma^2}\left(-\ln\left(\pi\sigma^2\right)\right) = -\frac{1}{\pi\sigma^2}\pi = -\frac{1}{\sigma^2} \\
&\frac{\partial}{\partial \sigma^2}\left(-\frac{\left|y_i - \Delta x_j\right|^2}{\sigma^2}\right) = -\left|y - \Delta x_j\right|^2 \left(-2\sigma^{-4}\right) = \frac{2\left|y_i - \Delta x_j\right|^2}{\sigma^4}
\end{aligned}$$

and solving $\sum_{i=1}^{n} \sum_{x_j \in x} Q_{X_i|Y_i}^{(t)}\left(x_j \mid y_i\right) \left[-\frac{1}{2} + \frac{\left|y_i - \Delta x_j\right|^2}{\sigma^2}\right] \stackrel{!}{=} 0$ easily for σ^2, we get (21).

[2]Page 10, here

Deriving (22) is a bit trickier than the others since for $p_($$\underline{x}$ we have a constrained optimization problem. We add this constraint into the cost function and form the Lagrangian function. Then we use Lagrangian for optimization. To optimize over p_j, we take the derivative of the Lagrangian with respect to p_j. Namely,

$$\frac{\partial}{\partial p_j} \sum_{i=1}^{n} \sum_{x_j \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)\left(-\ln\left(\pi\sigma^2\right) - \frac{\left|y_i - \Delta x_j\right|^2}{\sigma^2} + \ln p_j\right). + \lambda\left(1 - \sum_k p_k\right) \stackrel{!}{=} 0$$

$$\left(\sum_{i=1}^{n} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)\frac{1}{p_j}\right) - \lambda \stackrel{!}{=} 0 \Rightarrow p_j = \frac{\sum_{i=1}^{n} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)}{\lambda}$$

But we need to avoid complementary slackness variable λ, this can be done by

$$\sum_{x_j \in \mathcal{X}} p_j = \frac{\sum_{i=1}^{n}\sum_{x_j \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)}{\lambda} = 1$$

and since $\sum_{i=1}^{n}\sum_{x_j \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right) = n$, because $\sum_{x_j \in \mathcal{X}} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right) = 1$, then $\lambda = n$. Now, we can easily get rid of the λ by substituting n into it. Namely,

$$p_j = \frac{\sum_{i=1}^{n} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)}{\lambda} = \frac{\sum_{i=1}^{n} Q^{(t)}_{X_i|Y_i}\left(x_j \mid y_i\right)}{n}$$

Hence, (22) is also proved. After we derived the iterative algorithm for detecting the unknown parameters $\underline{\theta}$, we ran the K-means algorithm to select the initial distribution for $p(\underline{x})$ on the received symbols using the following K-means algorithm

E-step:

$$j^* = \underset{j \in \{1,2,\ldots,K\}}{\operatorname{argmin}} \left\|\underline{x}_j - \underline{y}_i\right\|^2.$$

M-step:

$$\frac{\partial}{\partial \underline{x}_j} \sum_{i=1}^{N}\sum_{j=1}^{K} \delta_{ji}\left\|\underline{x}_j - \delta_{ji}\underline{y}_i\right\|^2 = \sum_{i=1}^{N} 2\underline{x}_j\delta_{ji} - 2\delta_{ji}\underline{y}_i = 0$$

such that

$$\underline{x}_j = \frac{\sum_{i=1}^{N}\delta_{ji}\underline{y}_i}{\sum_{i=1}^{N}\delta_{ji}}$$

We also plotted the clusters and observed that M=4 is selected, and each constellation point is equally likely. Therefore, we select an initial guess of $p(\underline{x})$ as a uniform distribution, $\Delta = 1$, $\sigma^2 = 1$. And 200 of iterations, we estimate $\sigma^2 = 3.179$ and $\Delta = 3.315 + 0.458j$. And we also see that the initial guess of uniform distribution for $p(\underline{x})$ is the same as the original.

Please see the coding solutions for the graphs and solutions.