Solution to analysis in Home Assignment 1

QunZhang + qunz (cid)

Analysis

In this report I will present my independent analysis of the questions related to home assignment 1. I have discussed the solution with Lizi Teng, Mingxiang Zhao but I swear that the analysis written here are my own.

1 Properties of random variables

1.1 a

Since $X \sim N(\mu, \sigma^2)$, then the f(x) will be shown below:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

Then, assuming that $t = (x-\mu)/2\sigma$, the equation of the $E(x)$ will be shown below:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} (\mu + 2\sigma t) * \frac{2}{\sqrt{2 * \pi\sigma}} exp(-2t^2) * 2\sigma dt \tag{2}$$

$$= \int_{-\infty}^{\infty} \frac{2}{\sqrt{2 * \pi}} \mu * exp(-2t^2)dt + \int_{-\infty}^{\infty} \frac{4\sigma t}{\sqrt{2\pi}} exp(-2t^2)dt \tag{3}$$

Since t is an odd function, e^{-2t^2} is an even function,then te^{-2t^2} will be a odd function, then the equation could be shown below:

$$= \int_{-\infty}^{\infty} \frac{2}{\sqrt{2 * \pi}} \mu * exp(-2t^2)dt + \underbrace{\int_{-\infty}^{\infty} \frac{4\sigma t}{\sqrt{2\pi}} exp(-2t^2)dt}_{0} \tag{4}$$

$$= \frac{2}{\sqrt{2*\pi}}\mu \underbrace{\int_{-\infty}^{\infty} exp(-2t^2)dt}_{\frac{\sqrt{\pi}}{\sqrt{2}}} = \mu \tag{5}$$

Following the definition of the variance **Var[x]**:

$$Var[x] = E[(x-\mu)^2] = E(x^2) - 2E(x)\mu + \mu^2 = E(x^2) - \mu^2 \tag{6}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{\infty} (\mu + 2\sigma t)^2 * \frac{2}{\sqrt{2*\pi}\sigma} exp(-2t^2) * 2\sigma dt \tag{7}$$

$$= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu^2 exp(-2t^2) + 4\mu\sigma exp(-2t^2)t + 4\sigma^2 t^2 exp(-2t^2)dt \tag{8}$$

$$= \mu^2 * \frac{2}{\sqrt{2*\pi}} * \frac{\sqrt{\pi}}{\sqrt{2}} + 0 + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 4\sigma^2 t^2 exp(-2t^2)dt \tag{9}$$

$$\frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 4\sigma^2 t^2 exp(-2t^2)dt = -4\sigma^2 * \frac{2}{\sqrt{2\pi}} * \frac{1}{4} \int_{-\infty}^{\infty} td[exp(-2t^2)] \tag{10}$$

$$= -4\sigma^2 * \frac{2}{\sqrt{2\pi}} \{ \frac{1}{4} t * exp[-2t^2] \Big|_{-\infty}^{\infty} + \frac{1}{4} \int_{-\infty}^{\infty} exp[-2t^2]dt \} \tag{11}$$

$$= 0 + 4\sigma^2 * \frac{2}{\sqrt{2\pi}} * \frac{1}{4} * \frac{\sqrt{\pi}}{\sqrt{2}} = \sigma^2 \tag{12}$$

Thus, the equation of the variance is shown below:

$$Var[x] = \mu^2 + \sigma^2 - \mu^2 = \sigma^2 \tag{13}$$

1.2 b

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} Z * p(q)dq \\ &= \int_{-\infty}^{\infty} A * q * p(q)dq \\ &= A \underbrace{\int_{-\infty}^{\infty} q * p(q)dq}_{E(q)} = A * E(q) = A * \mu \end{aligned} \tag{14}$$

$$\begin{aligned} Cov(Z) &= E\{[Z-\mu][Z-\mu]^T\} \\ &= E\{[Aq - AE(q)][Aq - AE(q)]^T\} \\ &= E\{A[q-E(q)][q-E(q)]^T A^T\} \\ &= AE\{[q-\mu][q-\mu]^T\}A^T \\ &= ACov(q)A^T \end{aligned} \tag{15}$$

1.3 c

Since $Z = Aq$,where:

$$p(\mathbf{q}) = \mathcal{N}\left(\mathbf{q}; \begin{bmatrix} 0 \\ 10 \end{bmatrix}, \begin{bmatrix} 0.3 & 0 \\ 0 & 8 \end{bmatrix}\right), \mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \tag{16}$$

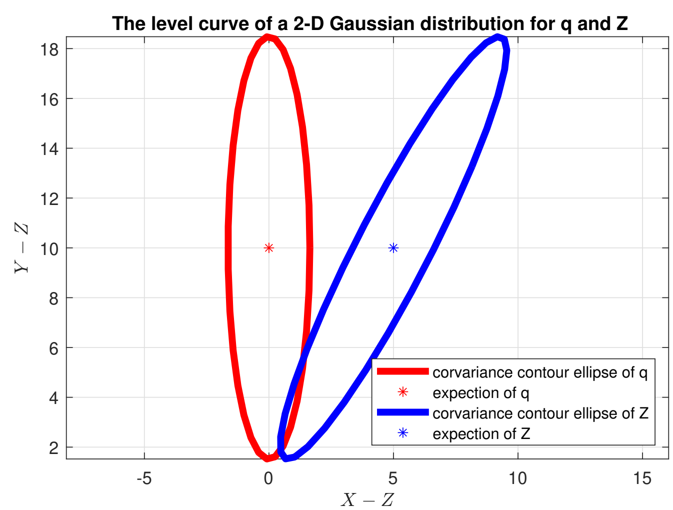


Figure 1: contour ellipse

As is shown above, which is the plot of ellipse of q and Z, the red one is the covariance contour ellipse of q and the blue one is the covariance contour ellipse of Z. Since matrix A is a transform matrix, which makes $\mu_z = A\mu_q$, μ_{z1} and μ_{z2} correlated, because $\mu_{z1} = \mu_{q1} + \mu_{q2}$ and $\mu_{z2} = \mu_{q2}$.
We can see that the A matrix shifts the mean by 5 in the x-direction, but

there is no change in the y-direction. As for the covariance matrix, the ellipse is scaled and rotated. Under the action of the A matrix, x and y are more related.So the ellipse is scaled and rotated.
The matrix **A** is also called an affine transformation matrix, which can be decomposed into a rotation matrix **R** and a scaling matrix **S**. Specifically, **A** can be decomposed as follows:

$$A = RS, S = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_1}, \end{bmatrix}, R = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \tag{17}$$

where λ_1 and λ_2 are the eigenvalues of Cov. The rotation matrix **R** can be computed from the eigenvectors of Cov and $\mathbf{v}_1$ and v_2 are the eigenvectors of Cov.
The rotation angle θ can be computed from the eigenvectors of Cov. Specifically, if $\mathbf{v}_1$ is the eigenvector corresponding to the larger eigenvalue, then θ is given by:

$$\theta = \tan^{-1}\left(\frac{v_{1,2}}{v_{1,1}}\right) \tag{18}$$

Then the $\mathbf{R}(\theta)$ could be expressed as the 2x2 rotation matrix that rotates counterclockwise by an angle θ:

$$R(\theta) = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{19}$$

2 Transformation of random variables

2.1 a

Taking $A = 3, b = 0$ into the **affineGaussianTransform** and **approxGaussianTransform**, we can get the analytical and numerical solutions respectively. As is shown in the figure 2, we can see that the analytical ($N(0, 18)$) and numerical solutions fit well.
Additionally, Because Z is a linear transformation relative to X, Z still satisfies the properties of Gaussian distribution, and different approximations can still make it fit into the form of Gaussian distribution.Since the variance of z is actually larger than the variance of X. Therefore, the distribution of

z is wider and flatter than the distribution of X. Additionally, the mean and variance transformed from X is necessary to describe its distribution.
When not so many values are sampled(fewer than 10000), the data fit will not be very good(as is shown in the figure 2, from 1000, 5000, 50000 to 100000). But when the sampled value is large enough so that the histogram does not change significantly, the numerical solution will tend to be more and more analytical. The histogram and true curve values will fit very well. To the different approximation, they fits well, and the μ and σ^2 tend to be 0 and 18, which fits better and better.

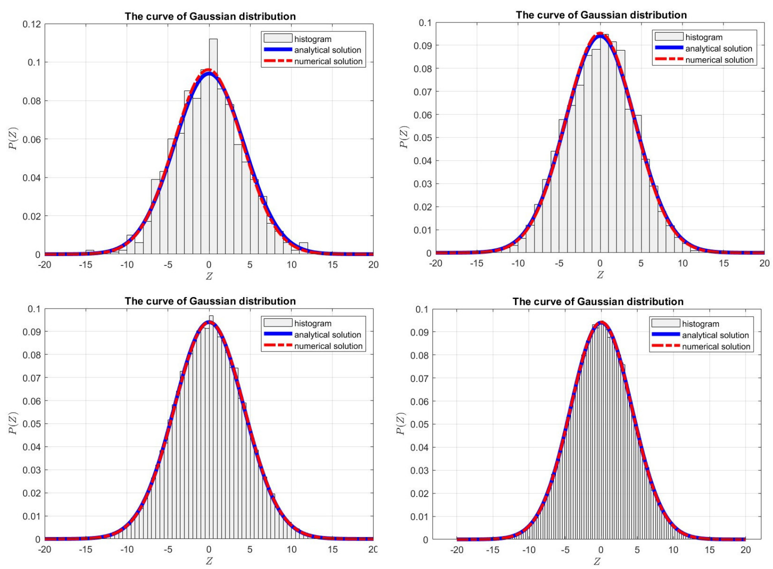


Figure 2: The curve of Gaussian distribution

2.2 b

According to the definition of the probability,the equation of $E(Z)$ are shown below:

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} q^3 p(q) dq \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} q^3 exp[-\frac{(q-\mu)^2}{2\sigma^2}] dq \end{aligned} \tag{20}$$

Since $x \sim N(0,2)$, and $q^3 * exp[-\frac{q^2}{4}]$ is an odd function,then the equation could be simplified as below:

$$\begin{aligned} E(x) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} q^3 exp[-\frac{q^2}{4}]dq \\ &= 0 \end{aligned} \tag{21}$$

The $Cov(Z)$ could be shown below (**calculated by the software Wolfram**):

$$\begin{aligned} Cov(Z) &= E(Z^2) - \mu^2 \\ &= \frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty} q^6 exp[-\frac{q^2}{4}]dq \\ &= \{120\sqrt{\pi}erf(\frac{q}{2}) + exp(-\frac{q^2}{4}(-2q^5 - 20q^3 - 120q))\} * \frac{1}{\sqrt{\pi}}\Big|_{-\infty}^{\infty} \\ &= \frac{1}{\sqrt{\pi}} * 120\sqrt{\pi} = 120 \end{aligned} \tag{22}$$

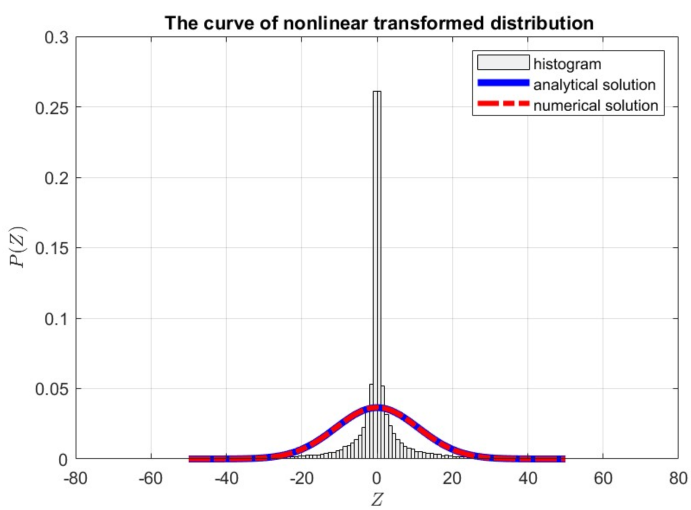


Figure 3: The curve of Gaussian distribution

As is shown in the figure 3, which are the curve of nonlinear transformed distribution, the histogram can not fit the curve of analytical solution and numerical solution. That is reasonable, because we tried to use the form of Gaussian distribution to describe it. However, due to the nonlinear transform, Z is not Gaussian distribution anymore, then they can not fit at all. If we have to get analytical solution, maybe more information maybe necessary, because only mean and variance are not necessary the distribution. Even through the sample go to 100000 til the histogram does not change, they can not fit.

2.3 c

If the transform is linear, the Z can also keep some properties of the X, such as same type of distribution. However, if going through a nonlinear transformation, its distribution will change.

3 Understanding the conditional density

3.1 a

According to the definition of the probability, the equation of $P(y)$ is shown below:

$$P(y) = \int_{-\infty}^{\infty} P(y|x) * p(x)dx \tag{23}$$

Since $y = H(x)+r$, then $P(y|x) = N(y; H(x), \sigma_r^2)$(when x, which means that $P(y|x)$ is determined. Thus, if we know the $p(x)$, for example, being defined such as uniform distribution, it can be determined.
However, since the $p(x)$ is unknown, the mixed distribution $P(y)$ can not be determined.

3.2 b

As what we mentioned above, since $P(y|x) = N(y; H(x), \sigma_r^2)$(when x is given, we can view it as a constant), we can determined that $p(y|x)$ is a Gaussian distribution without knowing the distribution of x.

$$\begin{aligned} p(y|x) &= \frac{1}{\sqrt{2\pi}\sigma_r} exp[-\frac{(y-H(x))^2}{2\sigma_r^2}] \\ &= N(y; H(x), \sigma_r^2) \end{aligned} \tag{24}$$

3.3 c

According to the equation(23), and we know the $H(x)$, if we know the distribution of $p(x)$, we can determine the distribution of the $p(y)$. Otherwise, we can not determine the $p(y)$. For the $p(y|x)$, we can determine it, $p(y|x) = N(y; H(x), \sigma_r^2)$.

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_r} exp[-\frac{(y-Hx)^2}{2\sigma_r^2}] \tag{25}$$

3.4 d

Since we know that $X \sim N(\mu_x, \sigma_x^2)$, we know the distribution of p(x) is Gaussian distribution, we can know the distribution of $p(y)$ and $p(y|x)$. $p(y|x)$ is also a Gaussian distribution. However, the distribution of $p(y)$ depends on the $H(x)$. If $H(x)$ is linear, we can know $p(y)$ is a Gaussian distribution.However, if $H(x)$ is none linear, it will not be a Gaussian distribution. If $H(x)$ is linear, then the distribution is shown below:

$$\begin{aligned} E(Y) = E(Hx + r) = Hx = \mu_y \\ E((x-\mu)^2) = H\sigma_x^2 H^T + \sigma_r^2 \\ Y \sim N(Hx, H\sigma_x^2 H^T + \sigma_r^2) \end{aligned} \tag{26}$$

If $H(x)$ is nonlinear, then the distribution is shown below:

$$\begin{aligned} p(y|x) &= \frac{1}{\sqrt{2\pi}\sigma_r} exp[-\frac{(y-H(x))^2}{2\sigma_r^2}] \\ &= N(y; H(x), \sigma_r^2) \end{aligned} \tag{27}$$

If we can determine the function of H(x), whether it is linear or nonlinear, we can also determine the probability distribution of p(y).

3.5 e

Assuming that $p(x)$ follows the distribution of the uniform distribution $U(a,b)$, $H(x) = 4x$(linear transformation). They the $P(y)$ could be shown below:

$$\begin{aligned} p(y) &= \int_a^b p(x) * \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{(y - H(x))}{2\sigma^2}]dx \\ &= -\frac{\sigma * erf(\frac{y-4x}{\sigma})}{8\sqrt{2}\sigma(b-a)}\Bigg|_a^b \end{aligned} \tag{28}$$

If the function is an nonlinear function, sadly we failed to express its analytical solution with a mathematical formula, but we can using the equation(20) and $Matlab$ to draw its analytical solution, which can be shown in the figure 5 (Assuming that $H(x) = x^2$).

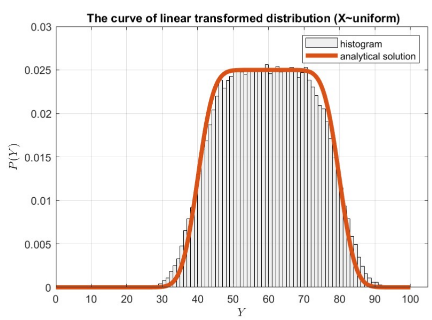


Figure 4: *linear transformed distribution*($X \sim U(a,b)$)

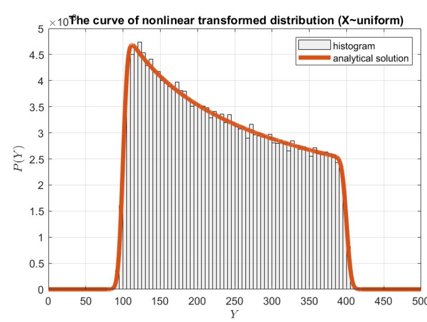


Figure 5: *nonlinear transformed distribution*($X \sim U(a,b)$)

Then, we view $p(x)$ as Gaussian distribution and $H(x)$ is a linear transformation (assuming $H(x) = 4x$).Since y will also be a Gaussian distribution($y = H(x) + r$), assuming that $X \sim (0, 25)$ then the analytical solution will be shown below:

$$\begin{aligned} E(Z) &= E(4X + r) = 4E(x) + E(r) \\ Cov(4x + r) &= 16Cov(x) + Cov(r) = 425 \end{aligned} \tag{29}$$

Then, the solution of histogram will be shown as Figure 6. However, if the $H(x)$ is nonlinear, then the distribution of Y will totally different. Using the same way in the *matlab*, then the plot of analytical solution and numerical solution will be shown in the figure 7.

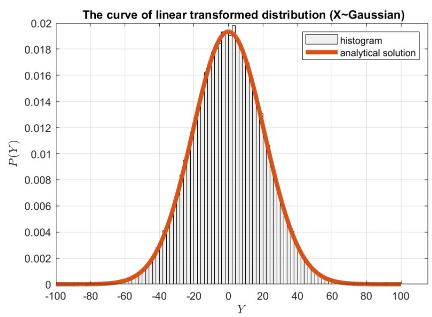


Figure 6: linear transformed distribution($X\ N(0, 25)$)

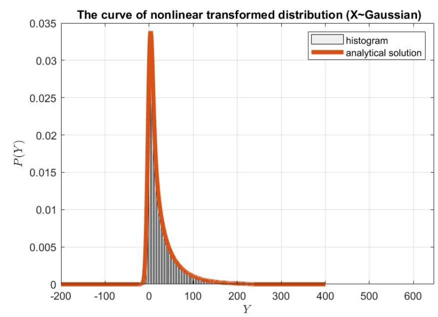


Figure 7: nonlinear transformed distribution($X\ N(0, 25)$)

Combining with the figure 4 and 6, we can see that different $p(x)$ will lead to different $p(y)$, **which verify that we can not determine the $p(y)$ without knowing $p(x)$**.
As what we mentioned in question 2, $p(y|x) = N(y; H(x), \sigma^2)$, then all of the distribution of $p(y|x)$ will be Gaussian distribution. Now that x is known, no matter what distribution it belongs to and what transformation H(x) belongs to, the final H(x) will be a scalar, so $p(y|x)$ will also be a Gaussian distribution, but the mean will be different, according to the figure, which verifies our theories before.

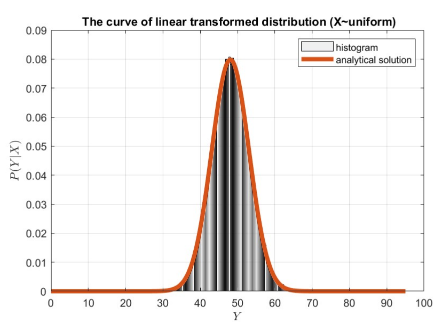


Figure 8: linear transformation with uniform distribution

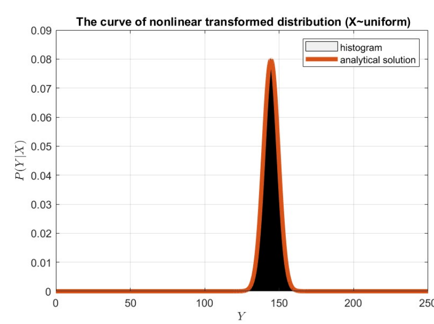
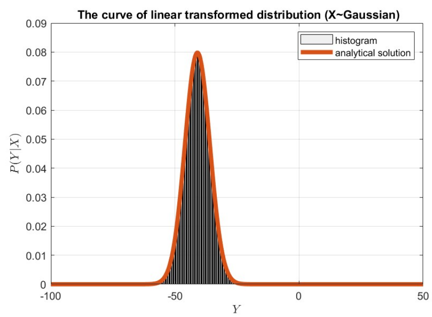


Figure 9: nonlinear transformation with uniform distribution



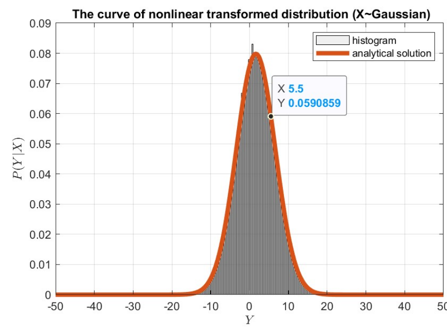


Figure 10: linear transformation with gaussian distribution Figure 11: nonlinear transformation with gaussian distribution

4 MMSE and MAP estimators

4.1 a

Through taking the distribution $N_1(\theta_1, \sigma^2)$ and $N_2(\theta_2, \sigma^2)$, where $\theta_1 = 1, \theta_2 = -1, \sigma^2 = 0.5^2$, then the histogram is shown as figure 12, which is a gaussian mixture model.

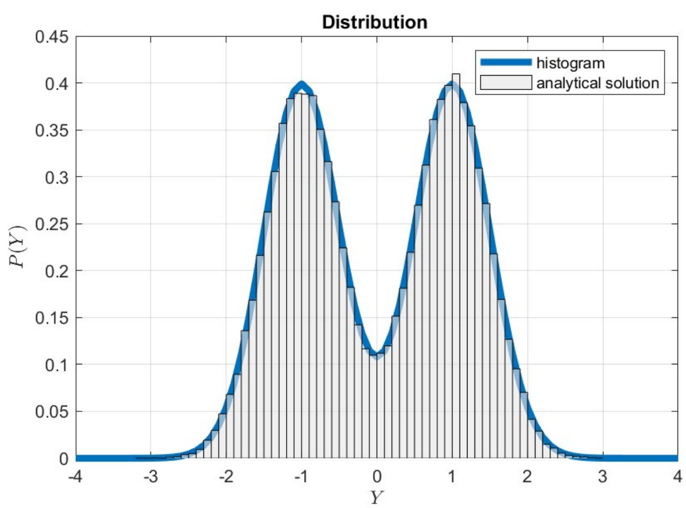


Figure 12: gaussian mixture model

4.2 b

$$p(\theta = 1|y = 0.7) = \frac{p(y = 0.7|\theta = 1) * p(\theta = 1)}{p(y = 0.7)} \tag{30}$$

$$p(\theta = -1|y = 0.7) = \frac{p(y = 0.7|\theta = -1) * p(\theta = -1)}{p(y = 0.7)} \tag{31}$$

Since $p(\theta = 1) = p(\theta = -1) = 0.5$, then we only consider the probability of $p(y = 0.7|\theta = -1)$ and $p(y = 0.7|\theta = 1)$. When $\theta = 1$, the probability $p(y = 0.7)$ is higher, and the probability is lower when $\theta = -1$, considering the distribution of $p(y) = N_1(-1, 0.5^2)$ and $p(y) = N_2(1, 0.5^2)$ respectively. So, I guess that $\theta = 1$.

4.3 c

Since $y = \theta + w$, then we can get that $p(y|\theta) = N(y; \theta, \sigma^2)$, the equation of $p(y)$ could be shown below:

$$\begin{aligned} p(y) &= \int_{\infty}^{\infty} p(y|\theta)p(\theta)d\theta \\ &= \sum p(y|\theta) \\ &= p(y|\theta_1) * p(\theta_1) + p(y|\theta_2) * p(\theta_2) \\ &= 0.5 * (N(y; \theta_1 = 1, \sigma^2) + N(y; \theta_2 = -1, \sigma^2)) \\ &= 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}} + 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}} \end{aligned} \tag{32}$$

4.4 d

$$\begin{aligned} p(\theta = 1|y) &= \frac{p(y|\theta = 1) * p(\theta = 1)}{p(y)} \\ &= \frac{0.5 * \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}}}{0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}} + 0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}} \\ &= \frac{e^{\frac{-(y-1)^2}{2\sigma^2}}}{e^{\frac{-(y-1)^2}{2\sigma^2}} + e^{\frac{-(y+1)^2}{2\sigma^2}}} = \frac{1}{e^{-\frac{2y}{\sigma^2}} + 1} \end{aligned} \tag{33}$$

$$
\begin{aligned}
p(\theta=-1|y) &= \frac{p(y|\theta=-1)*p(\theta=-1)}{p(y)} \\
&= \frac{0.5*\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}}{0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-1)^2}{2\sigma^2}}+0.5\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{e^{\frac{-(y+1)^2}{2\sigma^2}}}{e^{\frac{-(y-1)^2}{2\sigma^2}}+e^{\frac{-(y+1)^2}{2\sigma^2}}} = \frac{1}{e^{\frac{2y}{\sigma^2}}+1}
\end{aligned} \tag{34}
$$

4.5 e

$$
\begin{aligned}
\hat{\theta} &= \theta_1 * p(\theta_1 = 1|y) + \theta_2 * p(\theta_2 = -1|y) \\
&= \frac{e^{\frac{-(y-1)^2}{2\sigma^2}} - e^{\frac{-(y+1)^2}{2\sigma^2}}}{e^{\frac{-(y-1)^2}{2\sigma^2}} + e^{\frac{-(y+1)^2}{2\sigma^2}}} \\
&= \frac{e^{\frac{2y}{\sigma^2}}-1}{e^{\frac{2y}{\sigma^2}}+1} \\
&= tanh(\frac{y}{\sigma^2})
\end{aligned} \tag{35}
$$

4.6 f

To determine the optimal θ, we use the form of division. It should take $\theta = 1$ if the upper one is larger, and $\theta = -1$ if the lower one is larger, as described in the following formula:

$$
\begin{aligned}
\frac{p(\theta=1|y)}{p(\theta=-1|y)} &= e^{-\frac{(y-1)^2-(y+1)^2}{2\sigma^2}} \\
&= e^{\frac{4y}{2\sigma^2}} = e^{\frac{2y}{\sigma^2}}
\end{aligned} \tag{36}
$$

$$
\begin{cases}
\frac{p(\theta=1|y)}{p(\theta=-1|y)} \geq 1, \hat{\theta}_{MAP} = 1 & \text{if } y \geq 0 \\
\frac{p(\theta=1|y)}{p(\theta=-1|y)} < 1, \hat{\theta}_{MAP} = -1 & \text{if } y < 0
\end{cases} \tag{37}
$$

4.7 g

The plot of θ_{MMSE} and θ_{MAP} are shown below. From the plot we can see, if $y = 0.7$, according to the θ_{MAP}, the value of θ should be **1**. If taking

consideration of θ_{MMSE}, then the value should be 0.9926, which also means that we may have have choose to that $\theta = 1$. All the above conclusions have verified our guess, θ should be selected as $\theta = 1$.
In other words, MMSE is a measure of how well an estimator is expected to perform on average, while MAP is a method for finding the most likely estimate based on prior knowledge and the observed data.

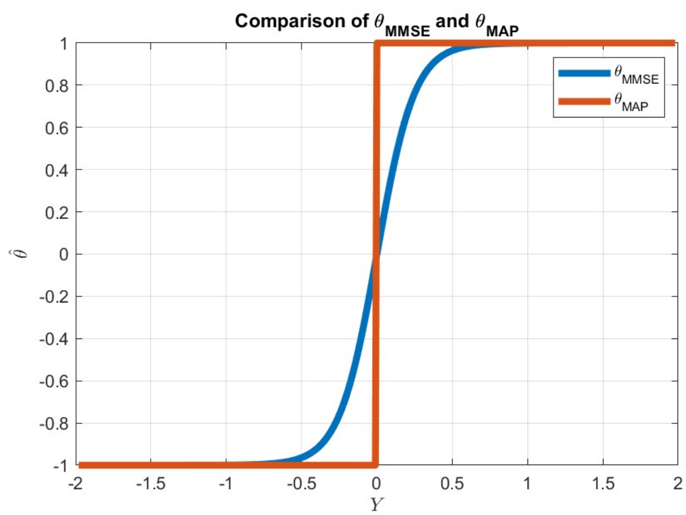


Figure 13: gaussian mixture model