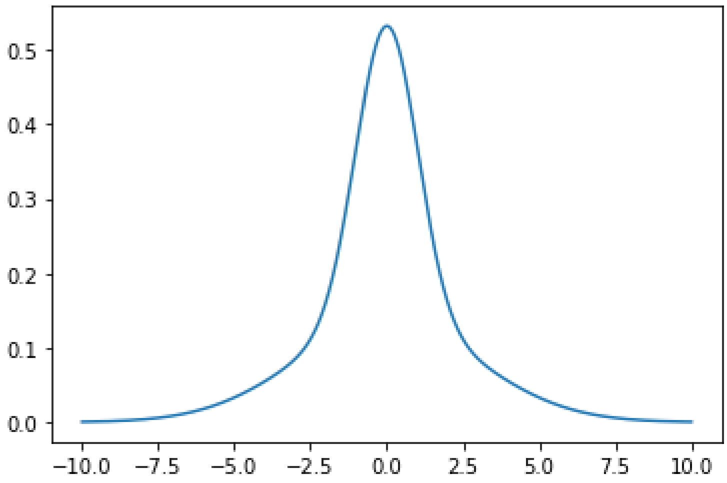
1.

```
import matplotlib.pyplot as plt
import scipy.stats
import numpy as np

x = np.linspace(-10, 10, 1000)
y1 = scipy.stats.norm.pdf(x, 0, 1)
y2 = scipy.stats.norm.pdf(x, 0, 3)

Y = y1 + y2
plt.plot(x, Y)
plt.show()
    --INSERT--
```



[5]

[2]

Image data in wavelet form could be modeled with a Gaussian scale mixture.

2.

The data likelihood is

$$
\begin{aligned}
p(X, Z|\theta) &= \prod_{i=1}^{n} p\left(z_i\right) p\left(x_i \mid z_i\right) \\
&= \prod_{i=1}^{n} \pi_{z_i} p\left(x_i \mid z_i\right) \\
&= \prod_{i=1}^{n} \prod_{j=1}^{K} \pi_j^{z_{ij}} p(x_i \mid z_i = j)^{z_{ij}}
\end{aligned}
$$

The complete data log-likelihood is

$$
\begin{aligned}
\log p(X, Z|\theta) &= \sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} \log \pi_j + z_{ij} \log p\left(x_i \mid z_i = j\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} \log \pi_j + z_{ij} \left[- \log \sqrt{2\pi}\sigma_j - \frac{1}{2}\left(\frac{x-\mu}{\sigma_j^2}\right)\right]
\end{aligned}
$$ [10]

3.

$$
\begin{aligned}
Q(\theta; \hat{\theta}^{\text{old}}) &= \mathbb{E}_{Z|X,\hat{\theta}^{\text{old}}}[\log p(X, Z \mid \theta)] \\
&= \mathbb{E}_{Z|X,\hat{\theta}^{\text{old}}}\left[\sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} \log \pi_j + z_{ij} \log p\left(x_i \mid z_i = j\right)\right] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{K} \hat{z}_{ij} \log \pi_j + \hat{z}_{ij} \left[- \log \sqrt{2\pi}\sigma_j - \frac{1}{2}\left(\frac{x_i-\mu}{\sigma_j^2}\right)^2\right]
\end{aligned}
$$ [5]

where

$$
\hat{z}_{ij} = p(z_i = j | x_i, \hat{\theta}^{\text{old}})
$$ [1]

4.

For general mixture models, the weights are updated as

$$
\hat{\pi}_j = \frac{\hat{n}_j}{n}
$$ [2]

where

$$
\hat{n}_j = \sum_{i=1}^{n} \hat{z}_{ij}
$$

For the variances,

$$
\sigma_j^* = \underset{\sigma_j}{\operatorname{argmax}}\, Q(\theta; \hat{\theta})
$$

The Q function containing σ_j is

$$
L_j = \sum_{i=1}^{n} \hat{z}_{ij} \left[- \log \sqrt{2\pi}\sigma_j - \frac{1}{2}\left(\frac{x_i-\mu}{\sigma_j^2}\right)^2\right]
$$

Take derivative and set to zero,

$$\frac{\partial L_j}{\partial \sigma j} = \sum_{i=1}^{n} \hat{z}_{ij} \left(-\frac{1}{\sigma_j} + \frac{(x_i - \mu)^2}{\sigma_j^3} \right) = 0$$

we get [3]

$$\sum_{i=1}^{n} \hat{z}_{ij}(x_i - \mu)^2$$

5.

The E-steps means the weights is the average samples that assigned to the component j. [3]

The M-steps means the variance is the variance weighted by $\hat{z}_{ij}$, the hidden values. [3]

EM for standard GMMS also involves the update of the mean, which is the sample average weighted by $\hat{z}_{ij}$, while EM for GMMSs with share mean does not. Besides this, they have the same form of parameter updating. [1]