COURSE: EIE6207 YEAR: 6
SUBJECT: Theoretical Fundamental and Engineering Approaches for Intelligent Signal and Information Processing

	SUBJECT EXAMINER	INTERNAL MODERATOR / ASSESSOR	EXTERNAL EXAMINER	
	M.W. Mak			

1. (a) (i)

$$L = ax^2 + by^2$$

(3 marks, K)

(ii) The constrained minimisation is given by

$$\begin{array}{ll} \min_{x,y} & ax^2 + by^2 \\ \text{subject to} & x + y = P \end{array}$$

Lagrangian is given by

$$\ell(x, y, \lambda) = ax^2 + by^2 + \lambda(P - x - y)$$

So, taking derivative of $\ell(x, y, \lambda)$ with respect to x, y, and λ, respectively, and set the results to 0, we obtain

$$\begin{aligned} \frac{\partial \ell}{\partial x} &= 2ax - \lambda = 0 \quad \Longrightarrow \quad x^* = \frac{\lambda}{2a} \\ \frac{\partial \ell}{\partial y} &= 2ay - \lambda = 0 \quad \Longrightarrow \quad y^* = \frac{\lambda}{2b} \\ \frac{\partial \ell}{\partial \lambda} &= P - x - y = 0 \quad \Longrightarrow \quad x^* + y^* = P \end{aligned}$$

Combining the 3 equations, we have

$$\lambda \left(\frac{1}{2a} + \frac{1}{2b} \right) = P \quad \Longrightarrow \quad \lambda = \frac{2Pab}{a + b}$$

This gives the result

$$x^* = \frac{Pb}{a + b} \quad \text{and} \quad y^* = \frac{Pa}{a + b}$$

(20 marks, A)

(iii) The minimum power loss is given by

$$L^* = a(x^*)^2 + b(y^*)^2 = \frac{ab^2P^2}{(a + b)^2} + \frac{a^2bP^2}{(a + b)^2} = \frac{abP^2}{a + b} = \alpha P^2$$

where we have defined $\alpha = \frac{ab}{a+b}$. Denote P_{old} and P_{new} as the old and new values of power transmission, respectively. Also, denote L^*_{old} and L^*_{new} as the old and new minimum power loss, respectively. If the power transmission is increased by 5%, we have

$$P_{\text{new}} = P_{\text{old}} + 0.05P_{\text{old}} = 1.05P_{\text{old}}.$$

Then, the percentage increase in power loss is

$$\frac{L^*_{\text{new}} - L^*_{\text{old}}}{L^*_{\text{old}}} = \frac{\alpha P^2_{\text{new}} - \alpha P^2_{\text{old}}}{\alpha P^2_{\text{old}}} = \frac{1.05^2 P^2_{\text{old}} - P^2_{\text{old}}}{P^2_{\text{old}}} = 0.1025,$$

which is equivalent to 10.25%.

(17 marks, E)

(b) (i) The latent variables of GMMs are the indicator variables indicating which of the Gaussians in the GMM is responsible for generating a data sample.

(5 marks, K)

(ii) If the data likelihood is directly maximized, there is no closed-form solutions for most of the probabilistic models. Without the closed-form solutions, stochastic gradient descent will be required. On the other hand, by introducing latent variables in the auxiliary function, close-form solutions can be obtained in the M-step (provided that the posterior expectations of the latent variables have been computed in the E-step) for most probabilistic models.

(5 marks, AE)

COURSE: EIE6207 YEAR: 6
SUBJECT: Theoretical Fundamental and Engineering Approaches for Intelligent Signal and Information Processing

	SUBJECT EXAMINER	INTERNAL MODERATOR / ASSESSOR	EXTERNAL EXAMINER	
	M.W. Mak			

2. (a) (i) Let $\mathbf{\Phi} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \cdots \ \boldsymbol{\phi}_M]$ and $\tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is the global mean of $\mathcal{X}$ and $M \leq D$. We enforce $\boldsymbol{\phi}_d^\mathsf{T}\boldsymbol{\phi}_d = 1$, $d = 1, \ldots, M$, by using a set of Lagrange multipliers $\{\lambda_d\}_{d=1}^M$:

$$\begin{aligned}
L(\mathbf{\Phi}, \{\mathbf{h}_i\}) &= \sum_{i=1}^N (\tilde{\mathbf{x}}_i - \mathbf{\Phi}\mathbf{h}_i)^\mathsf{T}(\tilde{\mathbf{x}}_i - \mathbf{\Phi}\mathbf{h}_i) + \sum_{d=1}^M \lambda_d(\boldsymbol{\phi}_d^\mathsf{T}\boldsymbol{\phi}_d - 1) \\
&= \sum_{i=1}^N (\tilde{\mathbf{x}}_i - \mathbf{\Phi}\mathbf{h}_i)^\mathsf{T}(\tilde{\mathbf{x}}_i - \mathbf{\Phi}\mathbf{h}_i) + \mathrm{tr}\{\mathbf{\Phi}\mathbf{\Lambda}_M\mathbf{\Phi}^\mathsf{T} - \mathbf{\Lambda}\} \\
&= \sum_{i=1}^N \mathbf{x}^\mathsf{T}\tilde{\mathbf{x}}_i - 2\mathbf{h}_i^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\tilde{\mathbf{x}}_i + \mathbf{h}_i^\mathsf{T}\mathbf{h}_i + \mathrm{tr}\{\mathbf{\Phi}\mathbf{\Lambda}_M\mathbf{\Phi}^\mathsf{T} - \mathbf{\Lambda}\}
\end{aligned}$$

where $\mathbf{h}_i \in \mathbb{R}^M$, $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_M, 0, \ldots, 0\} \in \mathbb{R}^{D\times D}$, $\mathbf{\Lambda}_M = \mathrm{diag}\{\lambda_1, \ldots, \lambda_M\} \in \mathbb{R}^{M\times M}$, and $\mathbf{\Phi} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \cdots \ \boldsymbol{\phi}_M] \in \mathbb{R}^{D\times M}$.

(8 marks, AE)

Setting $\frac{\partial L}{\partial \mathbf{\Phi}} = \mathbf{0}$ and $\frac{\partial L}{\partial \mathbf{h}_i} = \mathbf{0}$, we obtain:

$$\sum\nolimits_i \tilde{\mathbf{x}}_i\hat{\mathbf{h}}_i^\mathsf{T} = \hat{\mathbf{\Phi}}\mathbf{\Lambda}_M \quad \text{and} \quad \hat{\mathbf{\Phi}}^\mathsf{T}\tilde{\mathbf{x}}_i = \hat{\mathbf{h}}_i \implies \hat{\mathbf{h}}_i^\mathsf{T} = \tilde{\mathbf{x}}_i^\mathsf{T}\hat{\mathbf{\Phi}}$$

where we have used:

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}\{\mathbf{X}\mathbf{B}\mathbf{X}^\mathsf{T}\} = \mathbf{X}\mathbf{B}^\mathsf{T} + \mathbf{X}\mathbf{B} \quad \text{and} \quad \frac{\partial \mathbf{a}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^\mathsf{T}.$$

Therefore,

$$\sum\nolimits_i \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\mathsf{T}\hat{\mathbf{\Phi}} = \hat{\mathbf{\Phi}}\mathbf{\Lambda}_M \implies \mathbf{S}\hat{\mathbf{\Phi}} = \hat{\mathbf{\Phi}}\mathbf{\Lambda}_M$$

So, $\hat{\mathbf{\Phi}}$ comprises the M eigenvectors of

$$\mathbf{S} = \sum_{i=1}^N \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\mathsf{T} = \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\mathsf{T}.$$

(7 marks, KA)

(ii) This is because computing the covariance matrix $\sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\mathsf{T}$ and their eigenvectors are very expansive when D becomes very large.

(5 marks, K)

COURSE: EIE6207 YEAR: 6
SUBJECT: Theoretical Fundamental and Engineering Approaches for Intelligent Signal and Information Processing

	SUBJECT EXAMINER	INTERNAL MODERATOR / ASSESSOR	EXTERNAL EXAMINER	
	M.W. Mak			

(b) (i) $K = 4$

(3 marks, K)

(ii)

$$
\begin{aligned}
\frac{\partial E}{\partial w_{kj}^{(2)}} &= \frac{\partial E}{\partial a_k^{(2)}} \frac{\partial a_k^{(2)}}{\partial w_{kj}^{(2)}} \\
&= \frac{\partial E}{\partial o_k^{(2)}} \frac{\partial o_k^{(2)}}{\partial a_k^{(2)}} o_j^{(1)} \\
&= \left(o_k^{(2)} - t_k\right) \frac{\partial h(a_k^{(2)})}{\partial a_k^{(2)}} o_j^{(1)} \\
&= \left(o_k^{(2)} - t_k\right) h'(a_k^{(2)}) o_j^{(1)} \\
&= \left(o_k^{(2)} - t_k\right) h(a_k^{(2)}) \left(1 - h(a_k^{(2)})\right) o_j^{(1)},
\end{aligned}
$$

where we have used $h'(z) = h(z)(1 - h(z))$.

(7 marks, A)

(iii)

$$
w_{kj}^{(2)} \leftarrow w_{kj}^{(2)} - \eta \frac{\partial E}{\partial w_{kj}^{(2)}}
$$

(5 marks, K)

(c) (i) We assign the following variables to the Kalman filter:

$\mathbf{x}_t \leftarrow x_t$ Unobserved voltage from the power supply
$\mathbf{F}_t \leftarrow 1$ No change in the target output voltage over time
$\mathbf{w}_t \leftarrow w_t$ Noise of output voltage with variance of σ_v^2
$\mathbf{z}_t \leftarrow 1$ Measured voltage in Volt
$\mathbf{H}_t \leftarrow 1$ Voltmeter also gives Volt as output, same unit as v_t
$\mathbf{v}_t \leftarrow v_t$ measurement noise with variance of σ_m^2
$\mathbf{u}_t \leftarrow 0$ There is no control input
$\mathbf{R}_t \leftarrow \sigma_m^2$
$\mathbf{Q}_t \leftarrow \sigma_v^2$

So, the resulting Kalman filter becomes

$$x_t = x_{t-1} + w_t$$
$$z_t = x_t + v_t$$

where $w_t \sim \mathcal{N}(0, \sigma_v^2)$ and $v_t \sim \mathcal{N}(0, \sigma_m^2)$. (10 marks, E)

(ii) Yes, because the Kalman filter uses a recursive formula to update the state x_t, which essentially combines all of the measurements up to the current time step. Therefore, the longer we measures, the more measurements we have and x_t will become closer to the true output voltage when t increases. (5 marks, A)