Exercise for MA-INF 2201 Computer Vision WS23/24
19.12.2018
Submission until 14.01.2024
Christmas Special

1. **Convolution and Fourier Transform:**

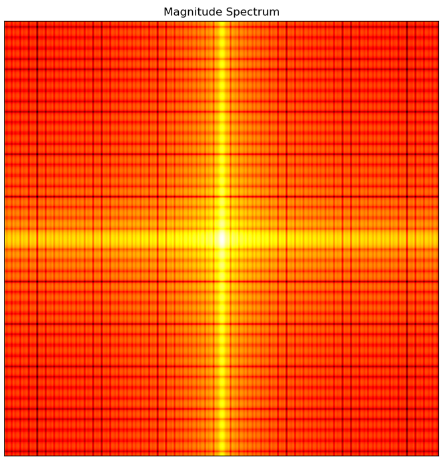
   (a) Compute the Fourier transform of the function

$$r(t) = \begin{cases} k, a/2 \leq t \leq 3a/2, \\ 0, \text{o.w.} \end{cases}$$

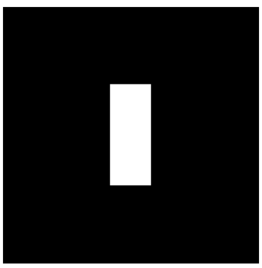
   (2 Points)
   Solution:

$$R(\omega) = \int_{-\infty}^{\infty} r(t)e^{-j\omega t}dt = \int_{a/2}^{3a/2} ke^{-j\omega t}dt = \frac{k}{-j\omega}\left[e^{-j\omega\frac{3a}{2}} - e^{-j\omega\frac{a}{2}}\right]$$
$$= \frac{2k}{\omega}e^{-j\omega a}\sin\left(\frac{\omega a}{2}\right) = e^{-j\omega a}\frac{ka\sin(\frac{\omega a}{2})}{\frac{\omega a}{2}} = e^{-j\omega a}ka\cdot\mathrm{sinc}\left(\frac{\omega a}{2}\right)$$

   (b) Below you see a frequency spectrum of an image.



   How did the original image look like? Explain why.
   (1.5 Points)
   Solution:



In the frequency spectrum, the oscillations of a sinc function can be observed in x- and y-direction, so the original function is rectangular. Since the oscillations are not identical in both directions, the image must be a rectangle. Because in the spectral domain sinc function lobes are wider along x-axis than along y-axis, the rectangle must be longer on the y-axis than on x-axis in the time domain.

(c) With which function does convolution keep the frequencies of a signal $s(t)$ (or image $I(x, y)$) unchanged? Why?
(1 Point)
Solution:
The desired function is the Dirac delta impuls $\delta(t) = \begin{cases} 1, t = 0 \\ 0, t \neq 0 \end{cases}$. Its Fourier transform is the constant 1. Convolution in the original (time or space) domain is a multiplication in the frequency domain and a multiplication by the constant 1 leaves the frequency spectrum unchanged.

(d) For the following 5×5 filter, determine if it separable. If yes, provide a separation. If not, argue why not.

$$A = \begin{pmatrix} 3 & 1 & -9 & -2 & 0 \\ 5 & 2 & 2 & 3 & -1 \\ 9 & 4 & -9 & -8 & 1 \\ 2 & 10 & -20 & -20 & 0 \\ 4 & 8 & 4 & -6 & 0 \end{pmatrix}$$

Solution:
A is not separable. One way to prove this is calculate $\text{rank}(A)$ and get $\text{rank}(A) > 1$. Alternatively, by definition, if A is separable then there $\exists$ two vectors $u, v \in \Re^5$ such that $vu^T = A$.
$u_1 v_1 = A_{11} \Rightarrow v_1 \neq 0$, $u_1 \neq 0$,
$v_1 u_5 = A_{15} \Rightarrow u_5 = 0$ but
$v_2 u_5 = A_{25} \neq 0 \Rightarrow u_5 = 0$, which is a contradiction, therefore no separation exists.

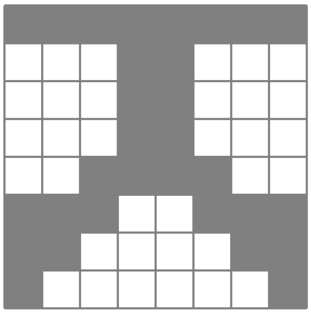
(e) For the following 5×5 filter, determine if it separable. If yes, provide a separation. If not, argue why not.

$$A = \begin{pmatrix} -21 & 6 & 3 & 12 & 9 \\ 7 & -2 & -1 & -4 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 35 & -10 & -5 & -20 & -15 \\ -14 & 4 & 2 & 8 & 6 \end{pmatrix}$$

(1 Point)
Solution:
It can be easily seen that the first column is a linear combination of the other columns, therefore $\text{rank}(A) = 1$ and the matrix is separable:

$$A = \begin{pmatrix} 3 \\ -1 \\ 0 \\ -5 \\ 2 \end{pmatrix} \begin{pmatrix} -7 & 2 & 1 & 4 & 3 \end{pmatrix}$$

(f) Compute the 2D distance transform of the below image by hand.



Provide the result after the initialization, after the forward pass, and after the backward pass.
(2.5 Points)
Solution:
Initialization:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \infty & \infty & \infty & 0 & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & 0 & 0 & \infty & \infty & \infty \\ \infty & \infty & \infty & 0 & 0 & \infty & \infty & \infty \\ \infty & \infty & 0 & 0 & 0 & 0 & \infty & \infty \\ 0 & 0 & 0 & \infty & \infty & 0 & 0 & 0 \\ 0 & 0 & \infty & \infty & \infty & \infty & 0 & 0 \\ 0 & \infty & \infty & \infty & \infty & \infty & \infty & 0 \end{pmatrix}$$

Forward and backward pass:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 2 & 2 & 2 & 0 & 0 & 1 & 2 & 2 \\ 3 & 3 & 3 & 0 & 0 & 1 & 2 & 3 \\ 4 & 4 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 3 & 3 & 2 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 2 & 2 & 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 3 & 3 & 2 & 1 & 0 \end{pmatrix}$$

2. **EM-Algorithm and Factor Analysis**: When working with images, a normal distribution with full covariance matrix is usually prohibitive as images are of very high dimension. A 100×100 pixel image already requires a $10,000 \times 10,000$ covariance matrix. Using a diagonal covariance matrix only can be a too strong limitation. *Factor analysis* provides a compromise by adding additional degrees of freedom to the model without using the full covariance matrix. Assuming D-dimensional observations, a matrix $\mathbf{\Phi} \in \mathbb{R}^{D \times K}(K \ll D)$ is used to extend the diagonal covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{D \times D}$. The final model then looks as follows:

$$Pr(x) = \mathcal{N}_x(\mu, \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{\Sigma}). \tag{1}$$

We define

$$Pr(x|h) = \mathcal{N}_x(\mu + \mathbf{\Phi}h, \mathbf{\Sigma}), \tag{2}$$
$$Pr(h) = \mathcal{N}_h(0, \mathbf{I}). \tag{3}$$

Then, Equation (1) can be rewritten as a marginalization by introducing a K-dimensional hidden variable h,

$$
\begin{aligned}
Pr(x) &= \int Pr(x|h)Pr(h)dh \\
&= \int \mathcal{N}_x(\mu + \mathbf{\Phi} h, \mathbf{\Sigma})\mathcal{N}_h(0, \mathbf{I})dh. \qquad (4)
\end{aligned}
$$

Note that Equation (1) and (4) are equivalent formulations of the same problem. Equation (4) allows us to optimize the model parameters using the EM-Algorithm.

(a) Given observations $x_1, \ldots, x_i, \ldots, x_I$, derive the E-Step of the EM-Algorithm for factor analysis, *i.e.* compute

$$\hat{q}_i(h_i) = Pr(h_i|x_i, \theta),$$

where $\theta = (\mu, \mathbf{\Phi}, \mathbf{\Sigma})$ denotes the set of model parameters. *Hint:* Terms that are independent of h_i are irrelevant later in the M-Step, so you can just represent them in a constant.
(2 Points)
Solution:
We first apply Bayes' rule to rewrite

$$\hat{q}_i(h_i) = \frac{Pr(x_i|h_i, \theta)Pr(h_i)}{Pr(x_i|\theta)} = \frac{\mathcal{N}_{x_i}(\mu + \mathbf{\Phi} h_i, \mathbf{\Sigma})\mathcal{N}_{h_i}(0, \mathbf{I})}{Pr(x_i|\theta)}.$$

Now, we apply the change of variables relation to the numerator and take $Pr(x_i|\theta)$ in the constant since it is independent of h_i,

$$\hat{q}_i(h_i) = \kappa_1 \mathcal{N}_{h_i}((\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}(x_i - \mu), (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi})^{-1})\mathcal{N}_{h_i}(0, \mathbf{I}).$$

The product of two normal distributions is again a normal distribution and we obtain

$$\hat{q}_i(h_i) = \kappa_2 \mathcal{N}_{h_i}((\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}(x_i - \mu), (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}).$$

The constant κ_2 can be omitted in order to obtain a proper probability distribution over h_i.

(b) Show that the update rules are

$$
\begin{aligned}
\tilde{\mu} &= \frac{1}{I}\sum_{i=1}^{I}\big(x_i - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i)\big), \\
\tilde{\mathbf{\Phi}} &= \Big(\sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\Big)\Big(\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\Big)^{-1}, \\
\tilde{\mathbf{\Sigma}} &= \frac{1}{I}\sum_{i=1}^{I}\operatorname{diag}\Big[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i)(x_i - \tilde{\mu})^T\Big].
\end{aligned}
$$

To make life easier for you, you may use that

$$
\begin{aligned}
&\arg\max_{\tilde{\theta}}\Big\{\sum_{i=1}^{I}\int \hat{q}_i(h_i)\log Pr(x_i, h_i|\tilde{\theta})dh_i\Big\} \\
=&\arg\max_{\tilde{\theta}}\Big\{\sum_{i=1}^{I}\mathbb{E}\big[-\log|\tilde{\mathbf{\Sigma}}| - (x_i - \tilde{\mu} - \tilde{\mathbf{\Phi}}h_i)^T\tilde{\mathbf{\Sigma}}^{-1}(x_i - \tilde{\mu} - \tilde{\mathbf{\Phi}}h_i)\big]\Big\}
\end{aligned}
$$

and

$$\mathbb{E}(h_i) = (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}(x_i - \mu),$$
$$\mathbb{E}(h_i h_i^T) = (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1} + \mathbb{E}(h_i)\mathbb{E}(h_i)^T.$$

(6 Points)

Solution:

We need to differentiate the log-likelihood function with respect to the parameters and set the derivative to zero. First, we expand the terms,

$$\begin{aligned}
L(\theta) &= \sum_{i=1}^{I}\mathbb{E}\big[-\log|\tilde{\mathbf{\Sigma}}| - (x_i - \tilde{\mu} - \tilde{\mathbf{\Phi}}h_i)^T\tilde{\mathbf{\Sigma}}^{-1}(x_i - \tilde{\mu} - \tilde{\mathbf{\Phi}}h_i)\big] \\
&= \sum_{i=1}^{I}\big[-\log|\tilde{\mathbf{\Sigma}}| - x_i^T\tilde{\mathbf{\Sigma}}^{-1}x_i - \tilde{\mu}_i^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mu}_i - \mathbb{E}(h_i^T\tilde{\mathbf{\Phi}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Phi}}h_i) \\
&\qquad + 2\tilde{\mu}^T\tilde{\mathbf{\Sigma}}^{-1}x_i - 2\tilde{\mu}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Phi}}\mathbb{E}(h_i) + 2x_i\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Phi}}\mathbb{E}(h_i)\big].
\end{aligned}$$

Now we take the derivatives with respect to the parameters and set them to zero.

$$\frac{\partial L}{\partial\tilde{\mu}} = \sum_{i=1}^{I} -2\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mu} + 2\tilde{\mathbf{\Sigma}}^{-1}x_i - 2\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Phi}}\mathbb{E}(h_i) \overset{!}{=} 0$$

Multiplication with $\tilde{\mathbf{\Sigma}}$ from the left and rearranging gives

$$\tilde{\mu} = \frac{1}{I}\sum_{i=1}^{I}\big(x_i - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i)\big).$$

The derivative with respect to $\tilde{\mathbf{\Phi}}$ is given by

$$\begin{aligned}
\frac{\partial L}{\partial\tilde{\mathbf{\Phi}}} &= \sum_{i=1}^{I} -2\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Phi}}\mathbb{E}(h_i h_i^T) - 2\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mu}\mathbb{E}(h_i)^T + 2\tilde{\mathbf{\Sigma}}^{-1}x_i\mathbb{E}(h_i)^T \overset{!}{=} 0 \\
\Rightarrow\quad \tilde{\mathbf{\Phi}}\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T) &= \sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T \\
\Rightarrow\quad \tilde{\mathbf{\Phi}} &= \Big[\sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\Big]\Big[\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\Big]^{-1}
\end{aligned}$$

The derivative with respect to $\tilde{\mathbf{\Sigma}}^{-1}$ is given by

$$\begin{aligned}
\frac{\partial L}{\partial\tilde{\mathbf{\Sigma}}^{-1}} &= \sum_{i=1}^{I}\operatorname{diag}\big[\tilde{\mathbf{\Sigma}} - x_i x_i^T - \tilde{\mu}\tilde{\mu}^T - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i h_i^T)\tilde{\mathbf{\Phi}}^T + 2x_i\tilde{\mu}^T - 2\tilde{\mu}\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T + 2x_i\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T\big] \\
&\overset{!}{=} 0
\end{aligned}$$

Multiplication with -1 and rearranging the terms yields

$$\begin{aligned} I \cdot \tilde{\mathbf{\Sigma}} &= \sum_{i=1}^{I} \mathrm{diag}\big[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T + \tilde{\mathbf{\Phi}}\mathbb{E}(h_i h_i^T)\tilde{\mathbf{\Phi}}^T - 2(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T\big] \\ &= \mathrm{diag}\big[\sum_{i=1}^{I}(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T + \tilde{\mathbf{\Phi}}\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\tilde{\mathbf{\Phi}}^T - 2\sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T\big] \end{aligned}$$

Plug in the results for $\tilde{\mathbf{\Phi}}$:

$$\begin{aligned} I \cdot \tilde{\mathbf{\Sigma}} = \mathrm{diag}\Big(&\sum_{i=1}^{I}(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T \\ &+ \Big[\sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\Big]\Big[\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\Big]^{-1}\Big[\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\Big]\tilde{\mathbf{\Phi}}^T \\ &- 2\sum_{i=1}^{I}(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T\big], \end{aligned}$$

which leads to the final result

$$\tilde{\mathbf{\Sigma}} = \frac{1}{I}\sum_{i=1}^{I}\mathrm{diag}\Big[(x_i - \tilde{\mu})(x_i - \tilde{\mu})^T - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i)(x_i - \tilde{\mu})^T\Big].$$

(Note that $\mathrm{diag}\big[\tilde{\mathbf{\Phi}}\mathbb{E}(h_i)(x_i - \tilde{\mu})^T\big] = \mathrm{diag}\big[(x_i - \tilde{\mu})\mathbb{E}(h_i)^T\tilde{\mathbf{\Phi}}^T\big]$.)

(c) What happens to the update rules if μ is initialized with the empirical mean, *i.e.* $\mu^{(0)} = \frac{1}{I}\sum_{i=1}^{I} x_i$?
(2 Points)
Solution:
Plug in $\mu^{(0)}$ into $\sum_i \mathbb{E}(h_i)$:

$$\begin{aligned} \sum_{i=1}^{I}\mathbb{E}(h_i) &= \sum_{i=1}^{I}(\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}(x_i - \frac{1}{I}\sum_{j=1}^{I}x_j) \\ &= (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\sum_{i=1}^{I}(x_i - \frac{1}{I}\sum_{j=1}^{I}x_j) \\ &= (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\Big(\sum_{i=1}^{I}x_i - \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{I}x_j\Big) \\ &= (\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\mathbf{\Phi} + \mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{\Sigma}^{-1}\Big(\sum_{i=1}^{I}x_i - \sum_{j=1}^{I}x_j\Big) \\ &= 0 \end{aligned}$$

So, for the update rule for $\tilde{\mu}$ we have

$$
\begin{aligned}
\tilde{\mu} &= \frac{1}{I}\sum_{i=1}^{I}\big(x_i - \tilde{\mathbf{\Phi}}\mathbb{E}(h_i)\big) \\
&= \frac{1}{I}\sum_{i=1}^{I} x_i - \frac{1}{I}\tilde{\mathbf{\Phi}}\sum_{i=1}^{I}\mathbb{E}(h_i) \\
&= \frac{1}{I}\sum_{i=1}^{I} x_i
\end{aligned}
$$

So, the dependence on $\tilde{\Phi}$ cancels and the mean does not need to be updated anymore but keeps its initial value.

(d) In order to start with a good initialization, one might want to initialize the model $\mathcal{N}_x(\mu, \mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{\Sigma})$ such that it is a normal distribution with diagonal covariance, *i.e.*

$$
\mu^{(0)} = \frac{1}{I}\sum_{i=1}^{I} x_i, \quad \mathbf{\Phi}^{(0)} = \mathbf{0}, \quad \mathbf{\Sigma}^{(0)} = \frac{1}{I}\sum_{i=1}^{I}\operatorname{diag}\big[(x_i-\mu)(x_i-\mu)^T\big].
$$

Is this beneficial? Why/why not?
(1 Point)

Solution:
This would be a very bad idea. If $\mathbf{\Phi}^{(0)} = \mathbf{0}$ then also $\mathbb{E}(h_i) = 0$. Thus, it is always true that $\tilde{\mathbf{\Phi}} = \mathbf{0}$ and $\tilde{\mu}$ and $\tilde{\mathbf{\Sigma}}$ also just keep their initial values and never change.