

Exercise for MA-INF 2201 Computer Vision WS18/19
14.12.2018
Submission until 5.01.2019
Christmas Special

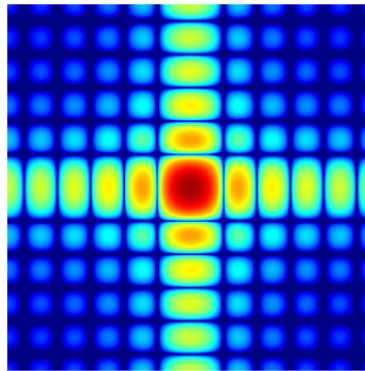
1. Convolution and Fourier Transform:

- (a) Compute the Fourier transform of the function

$$r(t) = \begin{cases} k, & |t| \leq a/2, \\ 0, & |t| > a/2. \end{cases}$$

(2 Points)

- (b) Below you see a frequency spectrum of an image. The frequency magnitudes are color coded, *i.e.* red is a high value, blue a low value.



How did the original image look like? Explain why.

(1.5 Points)

- (c) With which function does convolution keep the frequencies of a signal $s(t)$ (or image $I(x, y)$) unchanged? Why?

(1 Point)

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

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

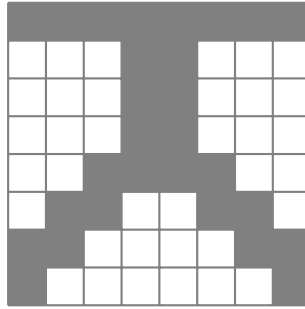
(1 Point)

- (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)

- (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)

2. **EM-Algorithm and Factor Analysis:** When working with images, a normal distribution with a full covariance matrix is usually prohibitive since 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 too strong a 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 $\Phi \in \mathbb{R}^{D \times K}$ ($K \ll D$) is used to extend the diagonal covariance matrix $\Sigma \in \mathbb{R}^{D \times D}$. The final model then looks as follows:

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

We define

$$Pr(x|h) = \mathcal{N}_x(\mu + \Phi h, \Sigma), \quad (2)$$

$$Pr(h) = \mathcal{N}_h(0, \mathbf{I}). \quad (3)$$

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

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

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 $x_1, \dots, x_i, \dots, x_I$, derive the E-Step of the EM-Algorithm for factor analysis, *i.e.* compute

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

where $\theta = (\mu, \Phi, \Sigma)$ 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)

- (b) Show that the update rules are

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

To make it easier, you may use that:

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

and

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

where \mathbb{E} is the expectation taken with respect to $\text{Pr}(h_i|x_i, \theta)$.

(6 Points)

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

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

Is this beneficial? Why/why not?

(1 Point)