Question 1: Image Features (14 pts.)

a) **Trajectory fitting** We wish to track the trajectory of a projectile being shot in the night sky. The projectile is equipped with a blinking flashlight, such that it becomes visible at regular intervals over its trajectory. We only have a single long-exposure gray-scale image of resolution 800x800 pixels to perform the task. Our image, like the one in Figure 1a, exposed the image sensor to the sky for five minutes, so that all the blinks from the projectile are visible at once on the image, and form a parabolic trajectory. However, also stars are visible on the background and are not clearly distinguishable from the blinks, if not for their random scattering in the image. Our objective is to find the parameters of the parabola traced by the flying projectile.

i) The first task is to clean all of the disturbance from the clouds and the background, and decide for every pixel whether it **could** belong to a blink or not. Figure 1b shows an example of the desired output of this step. Note that at this step stars don't need to be removed, but clouds and the general background should be removed. To study the statistics of the image in Figure 1a, we collect an intensity histogram of its pixels, reported in Figure 1c. Note that such histogram is reported with the vertical axis in **logarithmic scale**, to make small counts more visible. Briefly describe a procedure that could perform the task of cleaning the night image using the histogram. Provide explicit values for every parameter that you may use in your procedure. **2 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Since blinks are much brighter than the background, simple thresholding would suffice to identify them. However, this would not clean out stars, which are indistinguishable by just the intensity. Therefore given a pixel intensity value $0 \leq v \leq 255$ *it could be a blink if it satisfies* $v \geq t$ *for a threshold* t. *Looking at the histogram, we can observe a jump down by over an order of magnitude soon after the intensity value 50, which represents the higher intensity of the clouds in the background. Therefore, a threshold slightly larger than that would work (e.g.* $t \geq 60$*). To avoid cutting out some blinks we should not choose the threshold too high, therefore we should keep* $t \leq 150$ *to avoid cutting over the second strong count drop. The image in Figure 1b is produced with* $t = 100$.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

ii) We now have the binarized image in Figure 1b, where each pixel is either 0 (black, background) or 1 (white, a blink or a star). We want to further summarize this representation into a matrix $B \in \mathbb{R}^{N \times 2}$ where N is the number of blobs (i.e. clusters of white pixels connected together) in the binary image, and every row $b_i \in \mathbb{R}^2$ represents the average pixel coordinates of the i-th blob. Describe a procedure that would produce matrix B from the binarized image. **3 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Each blob is a connected region. Region growing can be used to identify all of the pixels belonging to the same blob. Therefore, blobs can be identified as connected components by region growing with a full 3x3 kernel, and for each identified connected component we can produce a row b_i *with the average of its pixels. The collection of vectors* b_i *forms the matrix* B.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

iii) Finally, we have the matrix B built in the previous exercise which lists N possible pixel locations where a blink could have occurred. Assuming that actual blinks are well aligned along a parabola, and stars are randomly scattered around the image, describe how to use a Hough transform to distinguish the blinks from the stars. In particular, you should describe:

- The dimensionality D of the necessary Hough transform.
- The parameterization of the necessary Hough transform.
- How a point $b_i \in \mathbb{R}^2$ is used to update the discrete Hough space.
- How to select the correct parameterization of the parabola.

5 pts.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

This problem can be formulated as a 3D Hough transform with a parametric parabola. Therefore:

- *The dimensionality D is $D = 3$*
- *A possible parameterization of the parabola is $y = ax^2 + bx + c$ with parameters $(a, b, c) \in \mathbb{R}^3$*
- *A point $b_i = (x_i, y_i)$ can be substituted into the parameterization, obtaining $y_i = ax_i^2 + bx_i + c$. This is a single linear equation with the three variables (a, b, c), therefore the solution has two degrees of freedom. Therefore, we can loop over all discrete values of two variables (for example a and b) and solve for the missing one ($c_{jk} = round(y_i - a_j x_i^2 - b_k x_i)$) to increase the count of the corresponding coordinates by one (in our example, at coordinates (a_j, b_k, c_{jk}) for every (i, j)). This should be repeated for every b_i in B.*
- *The point of the Hough space with the highest value is the parabola with the highest number of points going through that parabola, therefore it is our best candidate model.*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

b) **Correlation and convolution**

Let $K \in \mathbb{R}^{M \times M}$ be a kernel such that $K_{ij} \in \mathbb{R}$ and $i, j \in [-N, N] \subset \mathbb{Z}$, $M = 2N + 1$. We denote by $K * I$ the convolution of K on image I, and with $K \circ I$ the correlation of K on image I.

i) For which kernels K we have $K * I = K \circ I$ for every image I? Give an example of such kernel. **2 pts.**

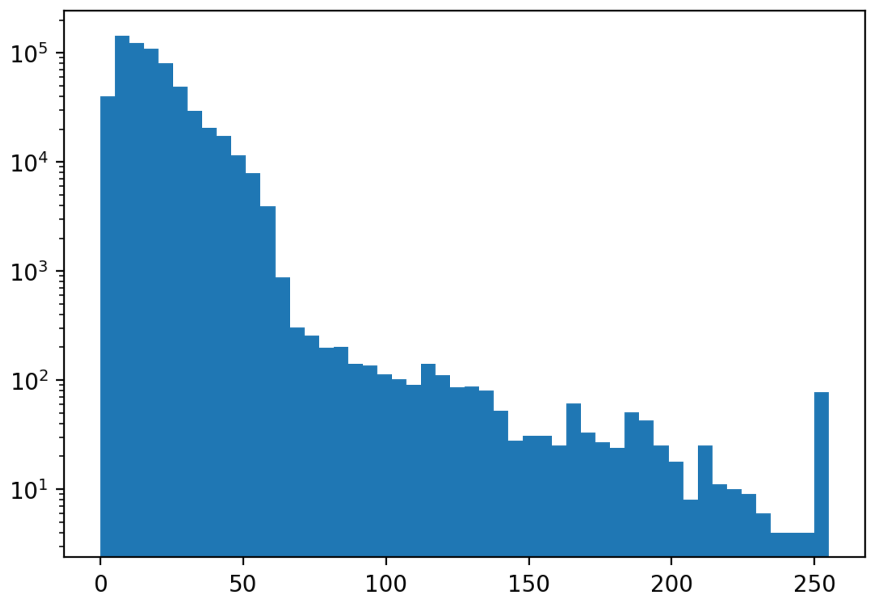
ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Convolution and correlation coincide when $K_{ij} = K_{-i-j}$, i.e. when the kernel is symmetric around its center. A Gaussian kernel is an example of symmetric kernel.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

(a) Original image of the night sky

(b) Binarized image of the night sky



(c) Logarithmic histogram of Figure 1a

Figure 1: A blinking projectile over the night sky.

ii) Let now K be defined by:

$$K_{ij} = \frac{1}{6\pi} e^{-\frac{i^2+j^2}{6}}$$

Name a use for this kernel, and show that it is separable. **2 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

This is a Gaussian smoothing kernel, particularly with $\sigma^2 = 3$. *It is used to perform isotropic Gaussian blurring. We can decompose* K *in:*

$$K_{ij} = \frac{1}{6\pi} e^{-\frac{i^2+j^2}{6}} = \frac{1}{\sqrt{6\pi}} e^{-\frac{i^2}{6}} \frac{1}{\sqrt{6\pi}} e^{-\frac{j^2}{6}}$$

Which leads to the definitions $F_i = \frac{1}{\sqrt{6\pi}} e^{-\frac{i^2}{6}}$, $G_j = \frac{1}{\sqrt{6\pi}} e^{-\frac{j^2}{6}}$

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

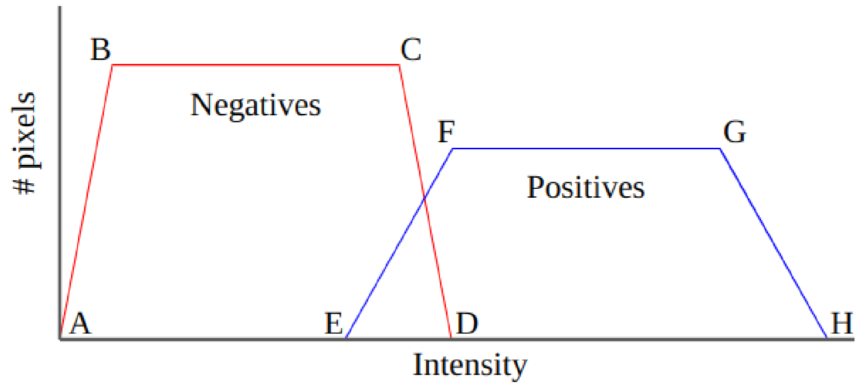


Figure 2: Separate histograms of Positives and Negatives. Refer to the text for coordinates of points A to H

Question 2: Basic image operations (9 pts.)

a) **Segmentation**

We wish to segment a grayscale image I with a simple thresholding approach. The intensity histogram of I separately for the positive class (in blue) and negative class (in red) is reported in Figure 2. The coordinates of points A to H are given: A(0, 0), B(20, 100), C(100, 100), D(120, 0), E(85, 0), F(120, 70), G(200, 70), H(235, 0). Points A to D define the histogram of the negative class, while points E to H define the histogram of the positive class. Let the threshold used be named τ:

i) What is the maximum True Positive Rate achievable? Give a value of τ that achieves it:

2 pts.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

$\tau = 0$ *achieves perfect True Positive Rate of 1. Since no positives are present until point E, values of* τ *smaller than 85 all maximize the True Positive Rate.*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

ii) What is the minimum False Positive Rate achievable? Give a value of τ that achieves it:

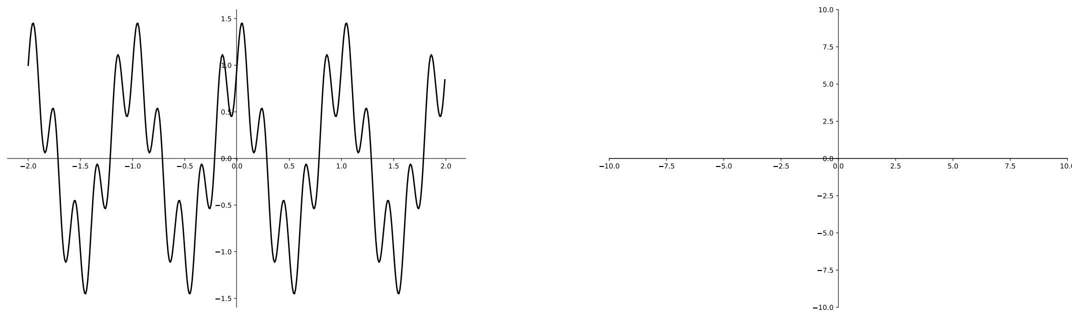
2 pts.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

$\tau = 255$ *achieves perfect False Positive Rate of 0. Since no negatives are present from point D, values of* τ *greater than 120 all minimize the False Positive Rate.*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

iii) What is the minimum number of incorrectly classified pixels? Give a value of τ that achieves it. Note: for this question you can assume that intensities and pixel counts are continuous quantities. **5 pts.**



(a) (b)

Figure 3: Superposition of sinusoidal signals and the corresponding Fourier transform.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

For every threshold τ, the number of misclassified pixels is the number of the positives smaller than τ plus the number of negatives higher than τ. This translates to the area of the Positives curve before τ plus the area of the Negatives after τ. In this case, this quantity is miminized at the intersection of the two curves and has value the area of the triangle with base ED and third vertex the intersection of EF with CD.

We need to find the intersection Q between lines EF and CD:

the equation of line EF: y = 2x - 170 the equation of line CD: y = -5x + 600

Their intersection is point Q with coordinates (110, 50).

Therefore the best threshold is $\tau = 110$ *which misclassifies* $(120 - 85) * 50/2 = 875$ *pixels*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Question 3: Fourier transform and filtering (12 pts.)

i) Figure 3a is showing a superposition of different sinusoidal signals. Draw the Fourier transform of this signal on Figure 3b. **2 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

The signal is the superposition of sinusoids of frequencies 1 and 5.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

ii) Consider the two images of Figure 5: which one has the highest frequency components? Justify your answer. **1 pts.**

ANSWER

Image 5b has the highest frequencies due to the dense foliage and diverse colors. On the contrary, Image 5a has fewer details and have rather low frequencies.

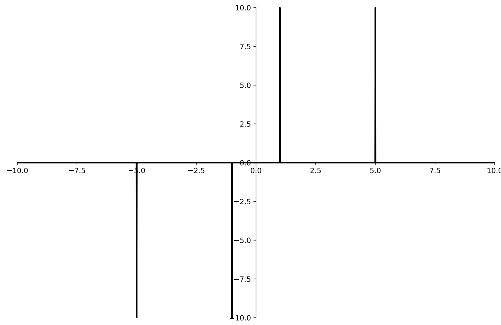


Figure 4: Fourier transform of the signal of Figure 3a

(a) (b)

Figure 5: Images with various frequencies.

ANSWER

iii) Suppose that we want to downsample Image 5b by a factor of 8 and that we know that the maximal frequency of the Fourier transform of the image is $\frac{1}{4}$. Is it possible to directly subsample the image without artifacts? If not, what problem could arise and how would you solve it? Justify all your answers. **4 pts.**

ANSWER

We cannot directly subsample the image without creating aliasing artifacts. Indeed, according to Nyquist theorem, the sampling frequency ($\frac{1}{8}$) should be at least twice the highest frequency in the signal ($\frac{1}{4}$), which is not the case here. One could overcome this issue by first convolving the image with a low-pass filter in order to remove too high frequencies (e.g. removing all frequencies higher than $\frac{1}{16}$).

ANSWER

iv) We now want to sharpen image 5a. What kind of filter would you use for that? Detail how you would apply it using the Fourier transform and frequency space. **3 pts.**

ANSWER

We can apply a high-pass filter to sharpen the image. To do so, we need to compute the Fourier transform of the image and the filter, and element-wise multiply both.

ANSWER

v) Among the following kernels, which one is the low-pass filter and which one the high-pass? **2 pts.**

$$(a)\begin{bmatrix}0.023 & 0.034 & 0.038 & 0.034 & 0.023\\ 0.034 & 0.049 & 0.056 & 0.049 & 0.034\\ 0.038 & 0.056 & 0.063 & 0.056 & 0.038\\ 0.034 & 0.049 & 0.056 & 0.049 & 0.034\\ 0.023 & 0.034 & 0.038 & 0.034 & 0.023\end{bmatrix} \quad (b)\begin{bmatrix}-0.0030 & -0.013 & -0.022 & -0.013 & -0.0030\\ -0.013 & -0.060 & -0.098 & -0.060 & -0.013\\ -0.022 & -0.098 & 0.84 & -0.098 & -0.022\\ -0.013 & -0.060 & -0.098 & -0.060 & -0.013\\ -0.0030 & -0.013 & -0.022 & -0.013 & -0.0030\end{bmatrix}$$

ANSWER

(a) is a low-pass filter (Gaussian kernel) and (b) is a high-pass filter (unit impulse minus a Gaussian kernel).

ANSWER

Question 4: Optical Flow (11 pts.)

a) The Lucas-Kanade algorithm can be used to estimate the optical flow of an image sequence. Nevertheless, it does not work for all cases. State the 3 conditions that have to be fulfilled so that this method works. **3 pts.**

ANSWER

1 point for each:

- *Brightness constancy: the intensity of the objects in the scene do not change in time.*
- *Small motion: objects move very slowly from frame to frame, which means that corresponding points from 2 consecutive images are not far apart.*
- *Spatial coherence: all points in a neighbourhood have the same motion.*

ANSWER

b) We consider a camera facing a planar scene along its Z axis, which points towards the scene. The camera captures images at a frame rate of 25 Hz and is a pinhole model with focal length $f = 500$ pixels. The camera plane is parallel to the scene and its projection equation relates a 3D point $\mathbf{P} = (X, Y, Z)$ to its pixel coordinates (x, y) by

$$x = \frac{fX}{Z}, \tag{1}$$

and similarly for y.

i) Express the 2D motion field $(\delta x, \delta y)$ (in pixels) of the 3D point $\mathbf{P}$ between two consecutive frames given its 3D velocity $\mathbf{V} = (V_x, V_y, V_z)$ in meters/second. Compute its value for $\mathbf{P} = (1, 2, 3)$ and $\mathbf{V} = (1, 0, 1)$. **5 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

- *We can write* $x = \frac{f\ X}{Z}$
- *Thus* $\frac{dx}{dt} = f\frac{ZV_x - XV_z}{Z^2}$ *(2 pts.)*
- *And* $\delta x = f\delta t\frac{ZV_x - XV_z}{Z^2}$ *(2 pts.)*
- *Numerical application:* $(\delta x, \delta y) = (\frac{40}{9}, -\frac{40}{9})$ *(1 pt.)*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

ii) What is the difference between this 2D motion field and the optical flow? **1 pt.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

0.5 point for each:

- *Motion field: projection of 3D scene flow*
- *Optical flow: apparent motion of the brightness pattern*

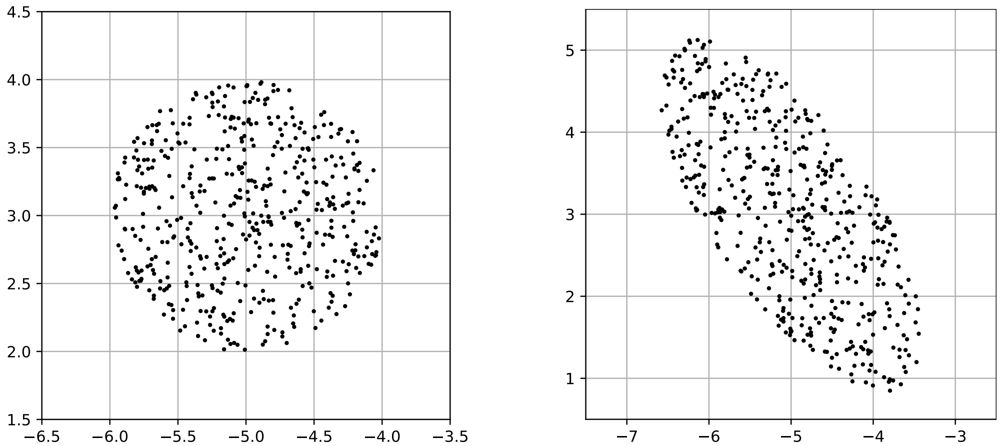
ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

iii) In which scenarios are the optical flow and the 2D motion field not equal? **2 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

When the brightness pattern is not uniquely constrained in all 2D directions: 1) texture-less area, 2) linear feature (aperture problem).

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER



(a) Distribution A. (b) Distribution B.

Figure 6: Illustrations for Question 5b.

Question 5: Principal Component Analysis (11 pts.)

a) We consider a dataset of 1000 faces images with size 40×40. Each array entry or number costs 1 unit of space. We wish to store this dataset given a budget of Z units of space. We choose to compress this dataset with PCA. Calculate the maximum number K of principal components allowed, taking into account all quantities required to decompress the dataset.

6 pts.

ANSWER

Let $x_i \in \mathbb{R}^{40\times40=1600}$ *be the original image and* $y_i \in \mathbb{R}^K$ *be a compressed image.*

- *Compression:* $y_i = (x_i - \mu)U_K$
- *Decompression:* $x_i = y_i U_K^\top + \mu$

We thus need to store:

- *the dataset mean* μ*:* $40 \times 40 = 1600$ ***1 pt.***
- *the truncated eigenmatrix* U_K*:* $1600 \times K$ ***1 pt.***
- *the compressed images* $\{y_i\}$*:* $1000 \times K$ ***1 pt.***

This yields

$$1600 + 1600 \times K + 1000 \times K \leq Z \quad \rightarrow \quad K \leq (Z - 1600)/2600$$

1.5 pts for setting the correct inequality, 1.5 pts for solving it correctly.

ANSWER

b) We consider two two-dimensional data distributions, A and B, shown in Figure 6.

i) Let us apply PCA to each set of point. In each plot, draw the expected sample mean and the two principal components. Clearly indicate which one is the largest, or explain when this is not possible. **4 pts.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

- *Correct means at* $(-5, 3)$ *for both.* ***1 pt.***
- *A: Explain that the two principal components are equal and orthogonal (1 pt) and that there is an infinite number of eigenvectors satisfying this (0.5 pt).* ***1.5 pt.***
- *B: Two correct principal components* $(-1, 1)$ *and* $(1, 1)$. ***1 pt.***
- *B: Correct largest component (the first one).* ***0.5 pt.***

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

ii) Given a new set of points, we wish to determine whether they were drawn from A or from B. Derive a simple test, with a unique threshold, based on the result of applying PCA to this point set. **1 pt.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Check the ratio of eigenvalues: if $\sigma_1/\sigma_2 > 1 + \epsilon$ *, then it is B, otherwise it is A. Other simpler solutions are also accepted as long as they clearly define how to obtain any threshold.*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER