1) Consider a point $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and a line passing through the origin which is represented by the vector $\mathbf{w} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$. What can you say about the following quantities? (MSQ)

(1) the projection of $\mathbf{x}$ onto the line

(2) the residue

The residue is equal to the zero vector.

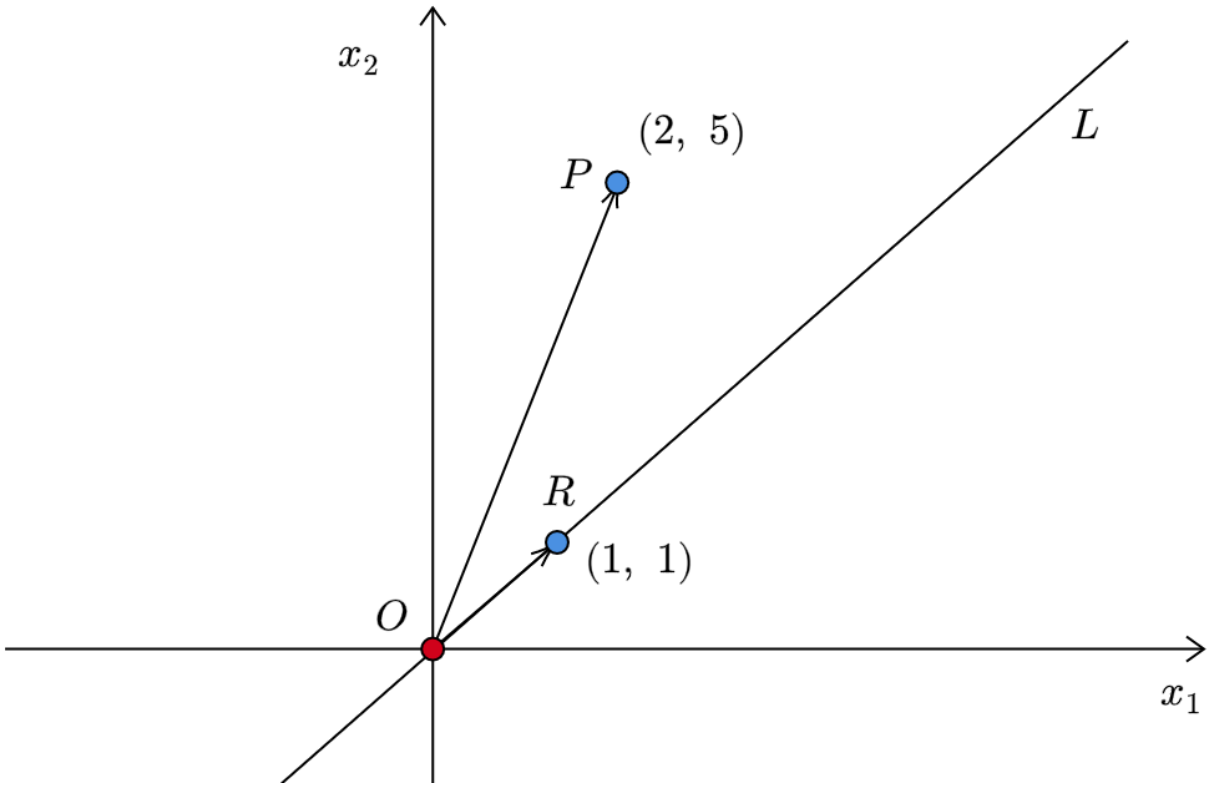
The residue is equal to the vector $\mathbf{x}$.

The projection is the zero vector.

The projection is equal to the vector $\mathbf{x}$.

Common data for questions (2) to (5)

Consider a point P and a line L that passes through the origin O. The point R lies on the line.



We use the following notation:

$\mathbf{w} = \overrightarrow{OR}$

$\mathbf{x} = \overrightarrow{OP}$

2) Consider the following statements:

Statement-1: The projection of $\mathbf{x}$ on the line L is given by $(\mathbf{x}^T\mathbf{w})\mathbf{w}$

Statement-2: The projection of $\mathbf{x}$ on the line L is given by $(\mathbf{x}^T\mathbf{w})\mathbf{x}$

Statement-3: The projection of $\mathbf{x}$ on the line L is given by $(\mathbf{x}^T\mathbf{x})\mathbf{w}$

Statement-4: The projection of $\mathbf{x}$ on the line L is given by $\mathbf{w}^T\mathbf{x}$

Which of the above statements is true?

- Statement-1
- Statement-2
- Statement-3
- Statement-4
- None of these statements are true.

3) Find the length of the projection of $\mathbf{x}$ on the line L. Enter your answer correct to two decimal places.

4.95

4) Find the residue after projecting $\mathbf{x}$ on the line L.

- $\begin{bmatrix} 3.5 \\ 3.5 \end{bmatrix}$
- $\begin{bmatrix} -1.5 \\ 1.5 \end{bmatrix}$
- $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$
- $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$

5) Find the reconstruction error for this point. Enter your answer correct to two decimal places.

4.5

6) Consider the following images of points in 2D space. The red line segments in one of the images represent the lengths of the residues after projecting the points on the line L. Which image is it?

Image-1:

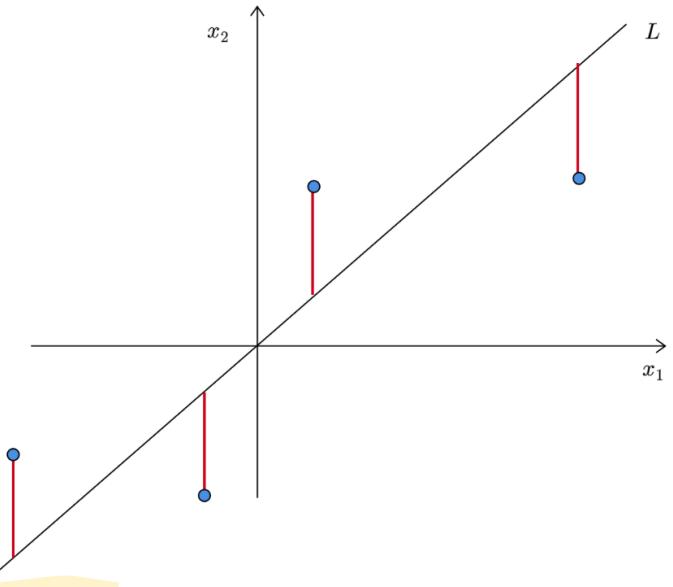
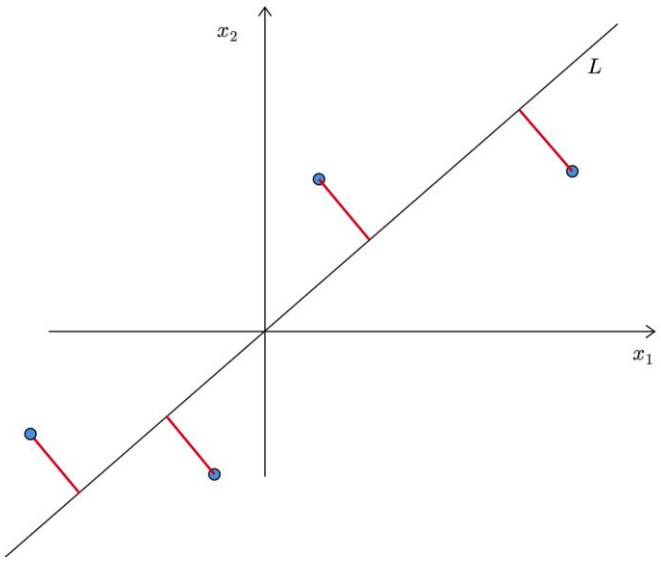


Image-2:

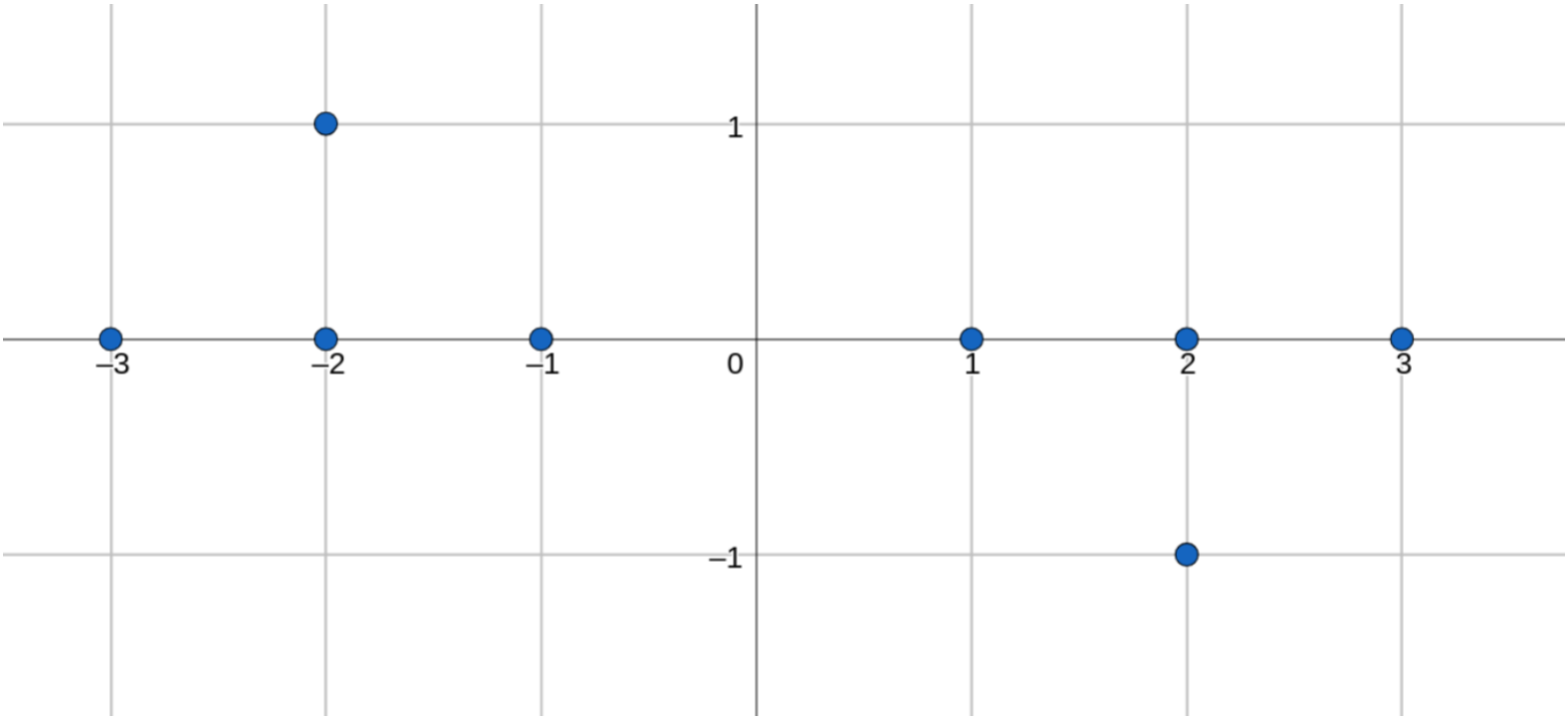


7) Consider a dataset that has 1000 samples, where each sample belongs to $\mathbb{R}^{30}$. PCA is run on this dataset and the top 4 principal components are retained, the rest being discarded. If it takes one unit of memory to store a real number, find the percentage decrease in storage space of the dataset by moving to its compressed representation. Enter your answer correct to two decimal places; it should lie in the range $[0, 100]$.

86.31

Common Data for questions (8) to (9)

Consider a dataset that has 8 points all of which belong to $\mathbb{R}^2$:



8) Find the covariance matrix of this dataset.

$\begin{bmatrix} 4.5 & -0.5 \\ -0.5 & 0.25 \end{bmatrix}$

$\begin{bmatrix} 36 & -4 \\ -4 & 2 \end{bmatrix}$

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

9) If PCA is run on this dataset, find the variance of the dataset along the first principal component. The eigenvectors of the covariance matrix are given below:

$\begin{bmatrix} -0.993 \\ 0.115 \end{bmatrix}$, $\begin{bmatrix} -0.115 \\ -0.993 \end{bmatrix}$

Recall that the first principal component is the most important. Enter your answer correct to two decimal places.

Note: Do not try to round off or drop decimal places in the intermediate computations. If a number has 4 places after the decimal, use all of them.

4.53

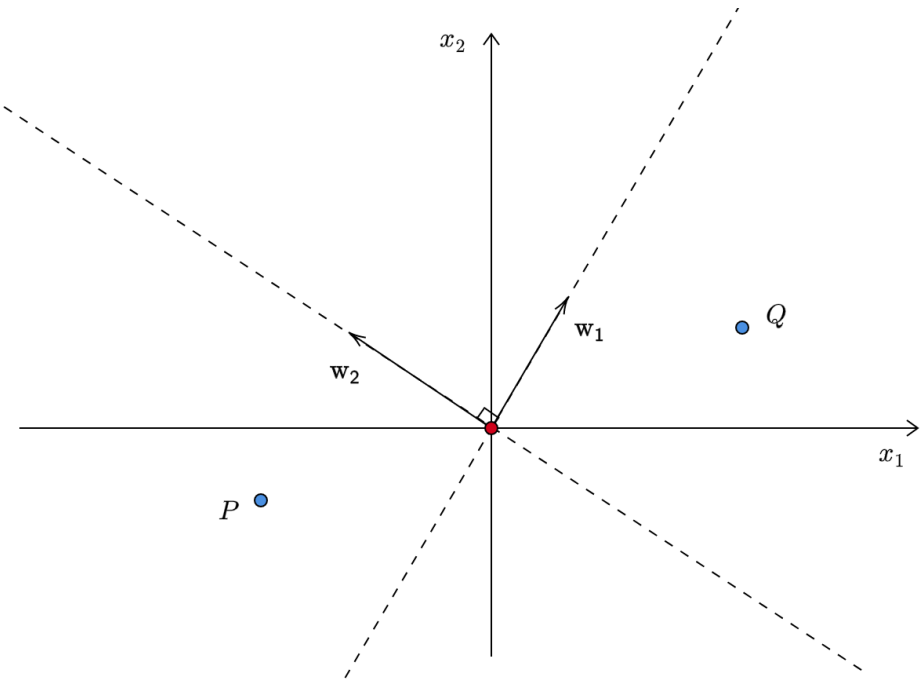
10) Consider a dataset of 100 points all of which lie in $\mathbb{R}^5$. The eigenvalues of the covariance matrix are given below:

$3.2, \quad 1.5, \quad 0.2, \quad 0.01, \quad 0.001$

If we run the PCA algorithm on this dataset and retain the top-k principal components, what is a good choice of k? Use the heuristic that was discussed in the lectures.

2

11) PCA is run on a dataset that has 2 features. The resulting principal components are $\mathbf{w}_1$ and $\mathbf{w}_2$. We represent the points in 2D space in terms of this new coordinate system made up of the principal components. The first coordinate corresponds to $\mathbf{w}_1$ and the second to $\mathbf{w}_2$. In such a scenario, what would be the sign of the coordinates for the points P and Q? ***1 point***



$P : (-ve, -ve)$

$P : (-ve, +ve)$

$Q : (+ve, +ve)$

$Q : (+ve, -ve)$