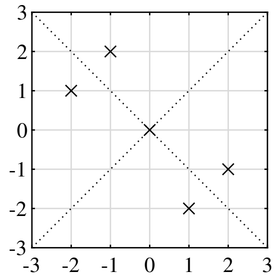
Exercises for Chapter 4

4.1 Let $X = \{(-2,1), (-1,2), (0,0), (1,-2), (2,-1)\}$.

a) Sketch a scatter diagram of this data set and the eigenvectors of its covariance matrix.



Because of symmetry, the mean of the data is at the origin. The symmetry axis along which the variance is maximum is the $-45°$ line. Any deviation from this line will reduce the variance, so the eigenvector with the maximum eigenvalue is in direction $(1, -1)$, and normalization yields $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$. Eigenvectors are pairwise orthogonal. Hence, if the first eigenvector is along the $-45°$ line, then the second eigenvector is along the $+45°$ line, which is in direction $(1, 1)$, and normalization yields $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. The direction of any of the eigenvectors may be reverted.

b) Compute the result of one-dimensional principal component projection.

Multiplication with the first eigenvector:
$(-2,1)^T \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = -\frac{3}{2}\sqrt{2}$
$(-1,2)^T \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = -\frac{3}{2}\sqrt{2}$
$(0,0)^T \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = 0$
$(1,-2)^T \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = \frac{3}{2}\sqrt{2}$
$(2,-1)^T \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = \frac{3}{2}\sqrt{2}$
$\Rightarrow \quad Y = \{-\frac{3}{2}\sqrt{2}, -\frac{3}{2}\sqrt{2}, 0, \frac{3}{2}\sqrt{2}, \frac{3}{2}\sqrt{2}\}$

c) Compute the average quadratic projection error.

Reverse projection of Y yields
$-\frac{3}{2}\sqrt{2} \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = (-\frac{3}{2}, \frac{3}{2})$
$0 \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = (0, 0)$
$\frac{3}{2}\sqrt{2} \cdot (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}) = (\frac{3}{2}, -\frac{3}{2})$
$\Rightarrow \quad X' = \{(-\frac{3}{2}, \frac{3}{2}), (-\frac{3}{2}, \frac{3}{2}), (0,0), (\frac{3}{2}, -\frac{3}{2}), (\frac{3}{2}, -\frac{3}{2})\}$.
The square distances between X and X' are
$\|(-2,1) - (-\frac{3}{2}, \frac{3}{2})\|^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$\|(-1,2) - (-\frac{3}{2}, \frac{3}{2})\|^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$\|(0,0) - (0,0)\|^2 = 0^2 + 0^2 = 0$

$\|(1,-2)-(\frac{3}{2},-\frac{3}{2})\|^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$
$\|(2,-1)-(\frac{3}{2},-\frac{3}{2})\|^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$
So, the average square distance is $e = \frac{1}{5}(\frac{1}{2}+\frac{1}{2}+0+\frac{1}{2}+\frac{1}{2}) = \frac{2}{5}$

This can also be done geometrically: First construct X' by mapping X to the $-45°$ line. Then look up the Euclidean distances between X and X'.

d) Sketch a Shepard diagram of this projection.

The pairwise distance matrices of X and Y are

$$D^X = \begin{pmatrix} 0 & \sqrt{2} & \sqrt{5} & 3\sqrt{2} & 2\sqrt{5} \\ \sqrt{2} & 0 & \sqrt{5} & 2\sqrt{5} & 3\sqrt{2} \\ \sqrt{5} & \sqrt{5} & 0 & \sqrt{5} & \sqrt{5} \\ 3\sqrt{2} & 2\sqrt{5} & \sqrt{5} & 0 & \sqrt{2} \\ 2\sqrt{5} & 3\sqrt{2} & \sqrt{5} & \sqrt{2} & 0 \end{pmatrix}$$

$$D^Y = \begin{pmatrix} 0 & 0 & \frac{3}{2}\sqrt{2} & 3\sqrt{2} & 3\sqrt{2} \\ 0 & 0 & \frac{3}{2}\sqrt{2} & 3\sqrt{2} & 3\sqrt{2} \\ \frac{3}{2}\sqrt{2} & \frac{3}{2}\sqrt{2} & 0 & \frac{3}{2}\sqrt{2} & \frac{3}{2}\sqrt{2} \\ 3\sqrt{2} & 3\sqrt{2} & \frac{3}{2}\sqrt{2} & 0 & 0 \\ 3\sqrt{2} & 3\sqrt{2} & \frac{3}{2}\sqrt{2} & 0 & 0 \end{pmatrix}$$

This yields a Shepard diagram with four points at $(\sqrt{2},0)$ (two-fold), $(\sqrt{5},\frac{3}{2}\sqrt{2})$ (four-fold), $(3\sqrt{2},3\sqrt{2})$ (two-fold), $(2\sqrt{5},3\sqrt{2})$ (two-fold).