Mathematics 327
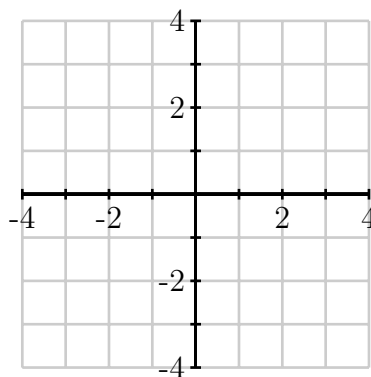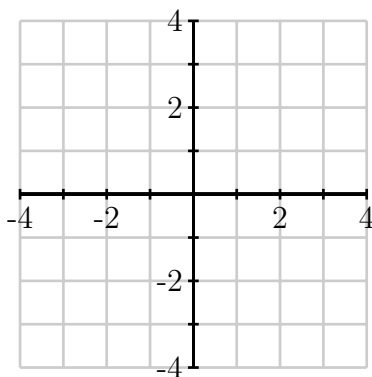**Principal component analysis**

1. Let's begin with the data set

$$(2, 1), (3, 0), (1, 3), (4, 0).$$

Plot the data below on the left

Now find the demeaned data and plot it above on the right.

In roughly what direction do you think the variance is greatest? Explain your thinking.

Construct the demeaned data matrix A and the covariance matrix C.

Find the variance in the direction $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and the variance in the direction $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
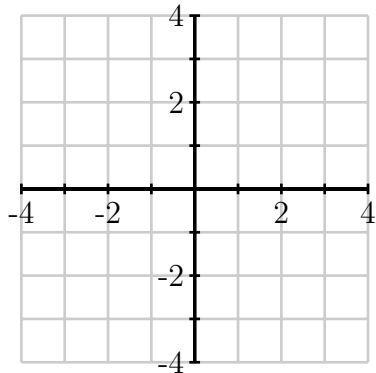
Find the direction of greatest variance. Sketch the demeaned data and the direction of greatest variance. What is the variance in this direction?
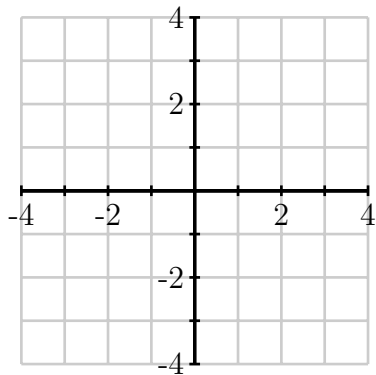
Find the orthogonal projections of the data points onto the direction of greatest variance. Sketch them above and find the resulting residual.

What is the total variance? What fraction of the variance is accounted for by the direction of greatest variance?

Find the direction of least variance. Sketch the demeaned data and the direction of least variance. What is the variance in this direction?

Find the orthogonal projections of the data points onto the direction of least variance. Sketch them above and find the resulting residual.

Suppose that your data points are in the form (x_1, x_2) and that you have another data point for which $x_1 = 3.5$. What is your prediction for x_2?