1. In class we showed that PCA finds the d-dimensional projection of $x_1, \ldots, x_n$ usch that the new components have maximal variance. Now show that the interpretation that PCA finds the linear projection to a d-dimensional subspace such that the reconstruction error is minimized is equivalent. Note: we found the reconstruction error for using only d principal components in class (for data with zero mean).

   Hint:

   Minimizing the reconstruction error is given by the least squares problem:

$$\min_{\substack{\mu,V,\beta \\ V^T V=I}} \sum_{k=1}^{n} \|x_k - (\mu + V\beta_k)\|_2^2.$$

   First find μ and β.

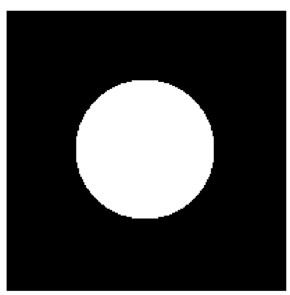
2. Plot a simple example in 2D for which calculating the SVD of X is not equivalent to calculating the Principal components of X. Explain.

3. Consider a $d \times n$ data matrix X, where $d < n$. Let $\{u_i\}_{i=1}^d$ be the d principle components of X, and let β_i equal,

$$\beta_i = u_i^T (X - \mu_s),$$

   where $\mu_s \in R^d$ is the sample mean vector. In class we referred to β_i as the new *meta features*. Prove that different meta features β_i, β_j have zero mean and are uncorrelated.

4. Coding problem: For the following image (circle.png), extract all overlapping patches of size 5×5 (for example using Matlab's im2col function or feature_extraction.image.extract_patches_2d in Python's scikit-learn).



   Calculate PCA of the resulting patch set. Plot the top 3 principal components and the corresponding principal directions (reshaped as patches of 5×5). Explain how the original image patches map onto the 3D coordinates and how this corresponds to the principal directions.

5. Compression. Load the file numbers.mat. The variable 'mat' is an image. Treat this image as a data matrix and apply PCA to it.

- Reconstruct the matrix $\hat{X}_d$ and plot it as a grayscale image using $d = 1, 10, 20, 100$ principal components (4 plots).
- Calculate the reconstruction error.
- Calculate the compression rate (using U_d, S_d, V_d instead of X).