with

$$U = \begin{pmatrix} -0.5286 & -0.6149 & 0.3017 & -0.5016 \\ -0.3803 & -0.3963 & -0.0933 & 0.8304 \\ -0.5520 & 0.2716 & -0.7603 & -0.2086 \\ -0.5209 & 0.6254 & 0.5676 & 0.1237 \end{pmatrix}, D = \begin{pmatrix} 30.2887 & 0 & 0 & 0 \\ 0 & 3.8581 & 0 & 0 \\ 0 & 0 & 0.8431 & 0 \\ 0 & 0 & 0 & 0.0102 \end{pmatrix}.$$

If we set $\sigma_3 = \sigma_4 = 0$, we obtain the best rank 2 approximation

$$A_2 = U(:, 1:2) * D(:, 1:2) * U(:, 1:2)' = \begin{pmatrix} 9.9207 & 7.0280 & 8.1923 & 6.8563 \\ 7.0280 & 4.9857 & 5.9419 & 5.0436 \\ 8.1923 & 5.9419 & 9.5122 & 9.3641 \\ 6.8563 & 5.0436 & 9.3641 & 9.7282 \end{pmatrix}.$$

A nice example of the use of Proposition 23.9 in image compression is given in Demmel [48], Chapter 3, Section 3.2.3, pages 113–115; see the Matlab demo.

Proposition 23.9 also holds for the Frobenius norm; see Problem 23.4.

An interesting topic that we have not addressed is the actual computation of an SVD. This is a very interesting but tricky subject. Most methods reduce the computation of an SVD to the diagonalization of a well-chosen symmetric matrix which is not $A^\top A$; see Problem 22.1 and Problem 22.3. Interested readers should read Section 5.4 of Demmel's excellent book [48], which contains an overview of most known methods and an extensive list of references.

23.4 Principal Components Analysis (PCA)

Suppose we have a set of data consisting of n points $X_1, \ldots, X_n$, with each $X_i \in \mathbb{R}^d$ *viewed as a row vector*. Think of the X_i's as persons, and if $X_i = (x_{i\,1}, \ldots, x_{i\,d})$, each $x_{i\,j}$ is the value of some *feature* (or *attribute*) of that person.

Example 23.5. For example, the X_i's could be mathematicians, $d = 2$, and the first component, $x_{i\,1}$, of X_i could be the year that X_i was born, and the second component, $x_{i\,2}$, the length of the beard of X_i in centimeters. Here is a small data set.