matrix norm. This concept is defined in Chapter 9, and the reader may want to review it before reading any further.

Given an $m \times n$ matrix of rank r, we would like to find a best approximation of A by a matrix B of rank $k \leq r$ (actually, $k < r$) such that $\|A - B\|_2$ (or $\|A - B\|_F$) is minimized. The following proposition is known as the *Eckart–Young theorem*.

Proposition 23.9. *Let A be an $m \times n$ matrix of rank r and let $VDU^\top = A$ be an SVD for A. Write u_i for the columns of U, v_i for the columns of V, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$ for the singular values of A ($p = \min(m, n)$). Then a matrix of rank $k < r$ closest to A (in the $\| \; \|_2$ norm) is given by*

$$A_k = \sum_{i=1}^{k} \sigma_i v_i u_i^\top = V \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) U^\top$$

and $\|A - A_k\|_2 = \sigma_{k+1}$.

Proof. By construction, A_k has rank k, and we have

$$\|A - A_k\|_2 = \left\| \sum_{i=k+1}^{p} \sigma_i v_i u_i^\top \right\|_2 = \left\| V \mathrm{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_p) U^\top \right\|_2 = \sigma_{k+1}.$$

It remains to show that $\|A - B\|_2 \geq \sigma_{k+1}$ for all rank k matrices B. Let B be any rank k matrix, so its kernel has dimension $n - k$. The subspace U_{k+1} spanned by $(u_1, \ldots, u_{k+1})$ has dimension $k + 1$, and because the sum of the dimensions of the kernel of B and of U_{k+1} is $(n - k) + k + 1 = n + 1$, these two subspaces must intersect in a subspace of dimension at least 1. Pick any unit vector h in $\mathrm{Ker}(B) \cap U_{k+1}$. Then since $Bh = 0$, and since U and V are isometries, we have

$$\|A - B\|_2^2 \geq \|(A - B)h\|_2^2 = \|Ah\|_2^2 = \left\|VDU^\top h\right\|_2^2 = \left\|DU^\top h\right\|_2^2 \geq \sigma_{k+1}^2 \left\|U^\top h\right\|_2^2 = \sigma_{k+1}^2,$$

which proves our claim. ∎

Note that A_k can be stored using $(m + n)k$ entries, as opposed to mn entries. When $k \ll m$, this is a substantial gain.

Example 23.4. Consider the badly conditioned symmetric matrix

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}$$

from Section 9.5. Since A is SPD, we have the SVD

$$A = UDU^\top,$$