Hermann Weyl. Sylvester's work was apparently "opaque." He gave a computational method to find an SVD. Schmidt's work really has to do with integral equations and symmetric and asymmetric kernels (1907). Weyl's work has to do with perturbation theory (1912). Autonne came up with the polar decomposition (1902, 1915). Eckart and Young extended SVD to rectangular matrices (1936, 1939).

**Theorem 22.5.** *(Singular value decomposition) For every real $n \times n$ matrix $A$ there are two orthogonal matrices $U$ and $V$ and a diagonal matrix $D$ such that $A = VDU^\top$, where $D$ is of the form*

$$D = \begin{pmatrix} \sigma_1 & & \dots & \\ & \sigma_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \sigma_n \end{pmatrix},$$

*where $\sigma_1, \dots, \sigma_r$ are the singular values of $A$, i.e., the (positive) square roots of the nonzero eigenvalues of $A^\top A$ and $A\,A^\top$, and $\sigma_{r+1} = \cdots = \sigma_n = 0$. The columns of $U$ are eigenvectors of $A^\top A$, and the columns of $V$ are eigenvectors of $A\,A^\top$.*

*Proof.* Since $A^\top A$ is a symmetric matrix, in fact, a positive semidefinite matrix, there exists an orthogonal matrix $U$ such that

$$A^\top A = UD^2U^\top,$$

with $D = \mathrm{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$, where $\sigma_1^2, \dots, \sigma_r^2$ are the nonzero eigenvalues of $A^\top A$, and where $r$ is the rank of $A$; that is, $\sigma_1, \dots, \sigma_r$ are the singular values of $A$. It follows that

$$U^\top A^\top AU = (AU)^\top AU = D^2,$$

and if we let $f_j$ be the $j$th column of $AU$ for $j = 1, \dots, n$, then we have

$$\langle f_i, f_j\rangle = \sigma_i^2\delta_{ij}, \quad 1 \leq i, j \leq r$$

and

$$f_j = 0, \quad r + 1 \leq j \leq n.$$

If we define $(v_1, \dots, v_r)$ by

$$v_j = \sigma_j^{-1} f_j, \quad 1 \leq j \leq r,$$

then we have

$$\langle v_i, v_j\rangle = \delta_{ij}, \quad 1 \leq i, j \leq r,$$

so complete $(v_1, \dots, v_r)$ into an orthonormal basis $(v_1, \dots, v_r, v_{r+1}, \dots, v_n)$ (for example, using Gram–Schmidt). Now since $f_j = \sigma_j v_j$ for $j = 1 \dots, r$, we have

$$\langle v_i, f_j\rangle = \sigma_j\langle v_i, v_j\rangle = \sigma_j\delta_{i,j}, \quad 1 \leq i \leq n,\ 1 \leq j \leq r$$