Generalizing the orthogonal diagonalization to any real matrix

We know that if $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix then there exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that $Q^{-1}AQ = Q^TAQ$ is diagonal.
This means: there is an appropriate orthonormal basis $\mathcal{B}$ such that the matrix of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ with respect to the basis $\mathcal{B}$ is diagonal.

We want to generalize this to any linear map:
Let $A \in \mathbb{R}^{m\times n}$. We want to find a pair of orthonormal bases $(\mathcal{B}, \mathcal{C})$ such that the matrix of the linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \mapsto A\mathbf{x}$ with respect to $(\mathcal{B}, \mathcal{C})$ is diagonal.
This means: we need to find an orthonormal basis $\mathcal{B}$ of $\mathbb{R}^n$ such that $\{ A\mathbf{b}_1, \dots, A\mathbf{b}_n \}$ is an orthogonal system of vectors.

Lemma. If $\mathbf{x}$ is an eigenvector of A^TA, and $\mathbf{x} \perp \mathbf{y}$, then $A\mathbf{x} \perp A\mathbf{y}$.
Proof. $\langle A\mathbf{y}, A\mathbf{x} \rangle = (A\mathbf{y})^T A\mathbf{x} = \mathbf{y}^T A^T A\mathbf{x} = \mathbf{y}^T \lambda \mathbf{x} = \lambda \langle \mathbf{y}, \mathbf{x} \rangle = 0.$

Corollary. We can choose $\mathcal{B}$ to be an orthonormal eigenbasis of A^TA (such a basis exists, since A^TA is symmetric). Then $A\mathbf{b}_1, \dots, A\mathbf{b}_n$ are orthogonal.
$\Rightarrow$ We can normalize and complete them to an orthonormal basis of $\mathbb{R}^m$.
(Actually, the eigenvalues are ≥ 0, so the diagonal form will have ≥ 0 diagonal elements.)

Singular Value Decomposition

Def. Let $A \in \mathbb{R}^{m\times n}$.
$A = U\Sigma V^T$ is a full singular value decomposition (**full SVD**) of A if
$U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and
$\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with diagonal elements $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$.
(So $U^{-1}AV = \Sigma$ is a diagonal matrix of $\mathbf{x} \mapsto A\mathbf{x}$, with respect to the pair $(\mathcal{B}, \mathcal{C})$ of bases, where $\mathcal{B}$ consists of the columns of V and $\mathcal{C}$ consists of the colums of U.)

Def.: The **singular values** of $A \in \mathbb{R}^{m\times n}$ are $\sigma_1 \geq \dots \geq \sigma_r > 0$, where $\sigma_1^2 \geq \ldots \geq \sigma_r^2 > 0$ are the positive eigenvalues of A^TA with multiplicities (the number of these is $r = \operatorname{rank} A^TA = \operatorname{rank} A$).

Def. Let $A \in \mathbb{R}^{m\times n}$ with $r = r(A)$.
$A = U_1\Sigma_1V_1^T$ is a reduced singular value decomposition (**reduced SVD**) of A if
$U_1 \in \mathbb{R}^{m\times r}$ and $V_1 \in \mathbb{R}^{n\times r}$ are semiorthogonal, and
$\Sigma_1 \in \mathbb{R}^{r\times r}$ is a diagonal matrix with diag. elements $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.

Theorem (Reduced SVD). *Every matrix $A \in \mathbb{R}^{m\times n}$ has a reduced SVD $A = U_1\Sigma_1 V_1^T$, where the diagonal elements of Σ_1 are the singular values of A.*

Proof: The proof also gives an algorithm for calculating the decomposition. Let the positive eigenvalues of the symmetric matrix A^TA be $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$, and $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ be an orthonormal system of eigenvectors for the given eigenvalues. Then $V_1 = [\mathbf{b}_1 \ldots \mathbf{b}_r]$ is a semiorthogonal matrix. Furthermore, $A\mathbf{b}_i \perp A\mathbf{b}_j$ for $i \neq j$ by the Lemma, and the lengthes of the vectors $A\mathbf{b}_i$ are $\sqrt{\mathbf{b}_i^TA^TA\mathbf{b}_i} = \sqrt{\lambda_i\mathbf{b}_i^T\mathbf{b}_i} = \sqrt{\lambda_i} = \sigma_i$. So $U_1 = AV_1\Sigma_1^{-1}$ is a semiorthogonal matrix. Hence $U_1\Sigma_1V_1^T = AV_1V_1^T$.

We only need to prove that $AV_1V_1^T = A$. We could complete the orthonormal system $\{\mathbf{b}_1, \ldots \mathbf{b}_r\}$ to a basis of $\mathbb{R}^n$ with eigenvectors $\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n$ of A^TA for the eigenvalue 0. Let $V = [\mathbf{b}_1 \ldots \mathbf{b}_n] = [V_1|T]$. Then V is invertible, and $AV_1V_1^TV = [AV_1V_1^TV_1|AV_1V_1^TT] = [AV_1|0] = A[V_1|T] = AV$, since $\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n \in \operatorname{Ker} A^TA = \operatorname{Ker} A$. We can simplify by V, and we get that $AV_1V_1^T = A$.

Example:

Pl.: $A = \begin{bmatrix} 1 & 2 \\ -2 & -4 \end{bmatrix}$, $A^TA = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}$, $k_{A^TA}(x) = x^2 - 25x = x(x-25)$, $\lambda_1 = 25$, $\lambda_2 = 0$, $\sigma_1 = 5$, $\Sigma = [5]$,

The eigenvector of A^TA for λ_1: $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, or to have an eigenvecor of length 1, it is $\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

$V_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\quad AV_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 5 \\ -10 \end{bmatrix}$,

We get U_1 by normalizing the columns of AV_1, in fact by dividing them by $\sigma_1, \ldots, \sigma_r$: $\quad U_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

$A = U_1\Sigma_1V_1^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot [5] \cdot \frac{1}{\sqrt{5}}[1 \quad 2]$.

Theorem (Full SVD). *Every matrix $A \in \mathbb{R}^{m\times n}$ has a full singular value decomposition.*

We can complete the U_1 and V_1 of the reduced SVD to orthogonal matrices: $V = [V_1|V']$ and $U = [U_1|U']$ and let Σ be the $m \times n$ diagonal matrix with the only nonzero diag. elements at the first r positions: $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Then

$$U\Sigma V^T = [U_1|U']\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ (V')^T \end{bmatrix} = [U_1\Sigma_1|\ 0]\begin{bmatrix} V_1^T \\ (V')^T \end{bmatrix} = U_1\Sigma_1V_1^T = A.$$

Example: Find the full SVD of the matrix of the previous example. $V = [V_1|V'] = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$, and $U = [U_1|U'] = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$.
Then $A = U\Sigma V^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$.

Applications of the SVD

Pseudoinverse: If $A = U_1\Sigma_1 V_1^T$ is a reduced SVD, then the pseudoinverse of A is $A^+ = V_1\Sigma_1^{-1}U_1^T$

Proof: It is easy to check the four properties of the pseudoinverse, using that $V_1^T V_1 = I_{r\times r}$ and $U_1^T U_1 = I_{r\times r}$.

Polar decomposition: If $A = PQ$ is a polar decomposition of the matrix $A \in \mathbb{R}^{n\times n}$ if P is a symmetric matrix with nonnegative eigenvalues, and Q is an orthogonal matrix.

Theorem. If $A = U\Sigma V^T$ is a full SVD, then $A = (U\Sigma U^T)(UV^T)$ is a polar decomposition.
Proof. $(U\Sigma U^T)^T = U\Sigma^T U^T = U\Sigma U^T$, and the eigenvalues of $U\Sigma U^T = U\Sigma U^{-1}$ are the same as the eigenvalues of Σ, which are its diagonal elements, so they are nonnegative.
$UV^T = UV^{-1}$ is a product of two orthogonal matrices, so it is also orthogonal.

Eckart–Young theorem about low-rank approximation. Let $A = U_1\Sigma_1 V_1^T$ be a reduced SVD, and $d < r = \operatorname{rank}(A)$. Then the best approximating matrix of rank at most d for A is

$$A^{(d)} = U^{(d)}\Sigma^{(d)}(V^{(d)})^T,$$

where $U^{(d)}$ and $V^{(d)}$ consist the first d columns of U_1 and V_1, respectively and $\Sigma^{(d)}$ is the left upper $d\times d$ submatrix of Σ_1. Here best approximating matrix means that $||A - M||$ is minimal among the matrices M with rank $M \le d$ if $M = A^{(d)}$, and for a matrix M, the norm of $||M||$ is $\sqrt{\sum_{i,j} m_{ij}^2}$.

Example. Determine the polar decomposition of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ -2 & -4 \end{bmatrix}.$$

Solution. Earlier we got the full SVD

$$A = U\Sigma V^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}.$$

So $A = PQ$ where

$$P = U\Sigma U^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$$

$$Q = UV^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ -4 & -3 \end{bmatrix}.$$

Example. Calculate the reduced SVD of the following matrix A. Use this to find the pseudoinverse and the best 1-rank approximation of A.

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -2 & 1 \end{bmatrix}$$

Solution.

$$A^T A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 8 & -2 \\ 0 & -2 & 1 \end{bmatrix}, \; k_{A^T A}(x) = -(x-9)(x-1)x \;\Rightarrow\; \begin{matrix} \lambda_1 = 9, & \lambda_2 = 1 \\ \sigma_1 = 3 & \sigma_2 = 1 \end{matrix}$$

Eigenvectors of $A^T A$:

$$\lambda_1 = 9: \quad \begin{bmatrix} -8 & 2 & 0 \\ 2 & -1 & -2 \\ 0 & -2 & -8 \end{bmatrix} \mapsto\mapsto \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{v}_1 = \frac{1}{3\sqrt{2}}\begin{bmatrix} -1 \\ -4 \\ 1 \end{bmatrix}$$

$$\lambda_1 = 1: \quad \begin{bmatrix} 0 & 2 & 0 \\ 2 & 7 & -2 \\ 0 & -2 & 0 \end{bmatrix} \mapsto\mapsto \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$V_1 = \frac{1}{3\sqrt{2}}\begin{bmatrix} -1 & 3 \\ -4 & 0 \\ 1 & 3 \end{bmatrix}, \quad AV_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -3 & 1 \\ 3 & 1 \end{bmatrix}, \quad U_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$A = U_1\Sigma_1 V_1^T = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{3\sqrt{2}}\begin{bmatrix} -1 & -4 & 1 \\ 3 & 0 & 3 \end{bmatrix}$$

$$A^+ = \frac{1}{3\sqrt{2}}\begin{bmatrix} -1 & 3 \\ -4 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & 4 \\ 2 & -2 \\ 4 & 5 \end{bmatrix}$$

$$A^{(1)} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}[3]\frac{1}{3\sqrt{2}}\begin{bmatrix} -1 & -4 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & 2 & -1/2 \\ -1/2 & -2 & 1/2 \end{bmatrix}$$