12.3. **Singular Value Decomposition.** The QR factorization plays an incredibly important role in solving systems of linear equations. Another, even more versatile factorization, is called the singular value decomposition. We develop this factorization while introducing the appropriate terminology.

Lemma 12.3.1. *Let $B \in \mathbb{R}^{m\times m}$ by symmetric and positive definite. The eigenvalues of B are real and positive, and the eigenvectors of B are orthogonal.*

Proof. Let α, v and β, w be two eigenvalue, eigenvector pairs with $\alpha \neq \beta$. Then,

$$\begin{aligned} \alpha\langle v, w\rangle = \langle \alpha v, w\rangle = \langle Bv, w\rangle = (Bv)^T w& \\ &= v^T B^T w = v^T(Bw) = \langle v, Bw\rangle = \langle v, \beta w\rangle = \beta\langle v, w\rangle. \\ \text{Therefore}\quad (\alpha-\beta)\langle v, w\rangle = 0.& \end{aligned}$$

Since $\alpha \neq \beta$, we must have $\langle v, w\rangle = 0$. (This only relies on the fact that B is symmetric.)

Since B is positive definite, $0 < x^T Bx$ for any $x \neq 0$. Therefore, since $Bv = \alpha v$,

$$0 < v^T Bv = v^T(\alpha v) = \alpha(v^T v) = \alpha\|v\|_2^2,$$

it must be that $\alpha > 0$. □

Lemma 12.3.2. *Let $A \in \mathbb{R}^{m\times n}$ with $m > n$. Then $A^T A \in \mathbb{R}^{n\times n}$ is symmetric and positive semi-definite. If A is full rank, then $A^T A$ is positive definite.*

Proof. $A^T A$ is symmetric since $(A^T A)^T = (A^T)(A^T)^T = A^T A$. Let $v \in \mathbb{R}^n$. Then,

$$0 \leq \|Av\|_2^2 = \langle Av, Av\rangle = (Av)^T(Av) = v^T(A^T A)v.$$

Thus, $A^T A$ is positive semi-definite.

Now $\|Av\|_2 = 0$ only if $Av = 0$. In this case, $(A^T A)v = 0$. If A is full rank, $A^T A$ is nonsingular. Thus, $v = 0$. Therefore, $A^T A$ is positive definite whenever A has full rank. □

Definition 12.1 Singular Values. Let $A \in \mathbb{R}^{m\times n}$ $(m > n)$ and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the real, nonnegative eigenvalues of $A^T A$. The *singular values* of A are the values

$$\sigma_i = \sqrt{\lambda_i} \quad \text{for } i = 1, \ldots, n.$$

Lemma 12.3.3. *The singular values σ_i of $A \in \mathbb{R}^{m\times n}$ satisfy*

$$\sigma_i = \|Av_i\|_2$$

where $A^T Av_i = \lambda_i v_i$ and $\|v_i\|_2 = 1$.

Proof. Homework. □

Theorem 12.3.1. *Suppose $\{v_1, \ldots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and $A^T Av_i = \lambda_i$ for $i = 1, \ldots, n$. Suppose A has r nonzero singular values. Then $\{Av_1, Av_2, \ldots, Av_r\}$ is an orthogonal basis for col(A).*

Proof. For $1 \leq i < j \leq n$,

$$\langle Av_i, Av_j\rangle = v_i^T(A^T A)v_j = \lambda_j v_i^T v_j = \lambda_j\langle v_i, v_j\rangle.$$

By, Lem. 12.3.1, the eigenvectors of $A^T A$ are orthogonal, thus $\langle Av_i, Av_j\rangle = 0$. Thus, $\{Av_1, \ldots, Av_n\}$ is an orthogonal set of vectors.

Clearly, $\text{span}\{Av_1, \ldots, Av_r\} \subset \text{col}(A)$. Let $x \in \mathbb{R}^n$ so that $Ax \in \text{col}(A)$. Since $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$, we have $x = \sum_{i=1}^n c_i v_i$. Therefore,

$$\begin{aligned} Ax = \sum_{i=1}^n c_i Av_i &= \sum_{i=1}^r c_i Av_i + \sum_{i=r+1}^n c_i Av_i \\ &= \sum_{i=1}^r c_i Av_i + 0 = \sum_{i=1}^r c_i Av_i \in \text{span}\{Av_1, \ldots, Av_r\}. \end{aligned}$$

Hence, $\text{col}(A) = \text{span}\{Av_1, \ldots, Av_r\}$. □

Theorem 12.3.2 (Singular Value Decomposition)**.** *Let $A \in \mathbb{R}^{m\times n}$, $m > n$, with r nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$. Then there exist matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, $\Sigma \in \mathbb{R}^{m\times n}$ such that U, V are orthogonal, Σ is diagonal with $\sigma_1, \ldots, \sigma_r$ the only nonzero values on the diagonal, and*

$$A = U\Sigma V^T.$$

Proof. Let $\{(\lambda_i, v_i)\}_{i=1}^r$ be the set of nonzero eigenvalue-eigenvector pairs of A^TA so that $\sigma_i = \|Av_i\|_2$ for $i = 1, \ldots, r$ and set $\sigma_j = 0$ for $j = r+1, \ldots, n$. Then, for $i = 1, \ldots, r$ define

$$u_i = \frac{1}{\|Av_i\|_2}Av_i = \frac{1}{\sigma_i}Av_i.$$

Then $\{u_1, \ldots, u_r\}$ is an orthonormal basis of $\text{col}(A) \subset \mathbb{R}^m$. Extend $\{u_1, \ldots, u_r\}$ to an orthonormal basis for $\mathbb{R}^m$ given by $\{u_1, \ldots, u_r, u_{r+1}, \ldots, u_m\}$. Let $\{w_{r+1}, \ldots, w_n\}$ be an orthonormal basis for $\text{null}(A)$. Now let

$$U = [u_1|u_2|\cdots|u_m] \in \mathbb{R}^{m\times m} \quad \text{and} \quad V = [v_1|v_2|\cdots|v_r|w_{r+1}|\cdots|w_n] \in \mathbb{R}^{n\times n}.$$

Let $D \in \mathbb{R}^{r\times r}$ be diagonal with $d_{ii} = \sigma_i$ for $i = 1, \ldots, r$ and define $\Sigma \in \mathbb{R}^{m\times n}$ by

$$\Sigma = \left[\begin{array}{c|c} D & 0^{r\times(n-r)} \\ \hline 0^{(m-r)\times r} & 0^{(m-r)\times(n-r)} \end{array}\right].$$

Then,

$$\begin{aligned}
U\Sigma &= [u_1|u_2|\cdots u_m]\left[\begin{array}{c|c} D & 0^{r\times(n-r)} \\ \hline 0^{(m-r)\times r} & 0^{(m-r)\times(n-r)} \end{array}\right] \\
&= [\sigma_1 u_1|\sigma_2 u_2|\cdots|\sigma_r u_r|0|0|\cdots|0] \\
&= [Av_1|Av_2|\cdots|Av_r|0|0|\cdots|0] \\
&= [Av_1|Av_2|\cdots|Av_r|Aw_{r+1}|Aw_{r+2}|\cdots|Aw_n] \\
&= AV.
\end{aligned}$$

By construction, U and V are orthogonal, so that

$$A = A(VV^T) = (AV)V^T = U\Sigma V^T.$$

□

Corollary 12.1 (Reduced SVD)**.** *Let $A \in \mathbb{R}^{m\times n}$, $m > n$, with r nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$. Then there exist matrices $\hat{U} \in \mathbb{R}^{m\times r}$, $\hat{V} \in \mathbb{R}^{n\times r}$, $\hat{\Sigma} \in \mathbb{R}^{r\times r}$ such that $\hat{U}^T\hat{U} = I_{r\times r} = \hat{V}^T\hat{V}$, $\hat{\Sigma}$ is diagonal with $\sigma_1, \ldots, \sigma_r$ along the diagonal, and*

$$A = \hat{U}\hat{\Sigma}\hat{V}^T.$$

Proof. Following the argument above, we have a full SVD for A, namely $A = U\Sigma V^T$. Let

$$\begin{aligned}
\hat{U} &= [u_1|u_2|\cdots u_r] \\
\hat{\Sigma} &= \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \\
\hat{V} &= [v_1|v_2|\cdots|v_r] \\
W &= [w_{r+1}|w_{r+2}|\cdots|w_n].
\end{aligned}$$

Then,

$$\begin{aligned}
A = U\Sigma V^T &= \left[\hat{U}\hat{\Sigma} \mid 0^{m\times(n-r)}\right]\begin{bmatrix} \hat{V}^T \\ W^T \end{bmatrix} \\
&= \hat{U}\hat{\Sigma}\hat{V}^T + 0^{m\times(n-r)}W^T = \hat{U}\hat{\Sigma}\hat{V}^T.
\end{aligned}$$

□

12.4. **Applications of the SVD.** Throughout the following discussion, let $A \in \mathbb{R}^{m\times n}$ have the singular value decomposition $A = U\Sigma V^T = \hat{U}\hat{\Sigma}\hat{V}$.

(A1): Rank: the rank of A is the number of nonzero singular values.

Proof. Since U, V are nonsingular, $\text{rank}(U) = \text{rank}(U^T) = m$, $\text{rank}(V) = \text{rank}(V^T) = n$. Also, if $B \in \mathbb{R}^{m\times n}$, then $\text{rank}(B) \le \min\{m, n\}$. Thus,

$$\text{rank}(A) \le \min\{\text{rank}(U^T), \text{rank}(V)\} \quad \text{and} \quad \text{rank}(\Sigma) \le \min\{\text{rank}(U), \text{rank}(V^T)\}.$$

Therefore

$$\begin{aligned}\text{rank}(A) = \text{rank}(U\Sigma V^T) &\le \min\{\text{rank}(U), \text{rank}(\Sigma), \text{rank}(V^T)\}\\ &= \text{rank}(\Sigma)\\ &= \text{rank}(U^T AV) \le \min\{\text{rank}(U^T), \text{rank}(A), \text{rank}(V)\} = \text{rank}(A).\end{aligned}$$

(A2): Determinants: If $m = n$ so that $A \in \mathbb{R}^{n\times n}$, then $\det(A) = \prod_{i=1}^{n} \sigma_i$.

Proof. Since U, V are orthogonal, $\det(U) = \det(V^T) = 1$.

$$\det(A) = \det(U\Sigma V^T) = \det(U)\det(\Sigma)\det(V^T) = \det(\Sigma) = \prod_{i=1}^{n} \sigma_i.$$

(A3): Inverses: If $m = n$ so that $A \in \mathbb{R}^{n\times n}$, then $A^{-1} = V\Sigma^{-1}U^T$.

(A4): Least Squares: If A is full rank, the least squares solution to $Ax = b$ is $\bar{x} = \hat{V}\hat{\Sigma}^{-1}\hat{U}^T b$.

Proof. We know that $\bar{x}$ is the solution to the normal equations $A^T A\bar{x} = A^T b$. Note that since A is full rank then $r = n$. Hence $\hat{V} = V \in \mathbb{R}^{n\times n}$ is orthogonal. Therefore,

$$A^T = V\hat{\Sigma}\hat{U}^T \quad \text{and} \quad A^T A = (V\hat{\Sigma}\hat{U}^T)(\hat{U}\hat{\Sigma}V^T) = V\hat{\Sigma}^2 V^T.$$

Thus, $A^T A$ is invertible and

$$\begin{aligned}\bar{x} = (A^T A)^{-1}A^T b &= (V\hat{\Sigma}^{-2}V^T)(V\hat{\Sigma}\hat{U}^T)b\\ &= (V\hat{\Sigma}^{-2})(V^T V)(\hat{\Sigma}\hat{U}^T)b\\ &= V(\hat{\Sigma}^{-2}\hat{\Sigma})\hat{U}^T b = V\hat{\Sigma}^{-1}\hat{U}^T b.\end{aligned}$$

(A5): Pseudo Inverse: The pseudo-inverse of a full rank matrix $A \in \mathbb{R}^{m\times n}$ is the matrix

$$A^\dagger = (A^T A)^{-1}A^T = V\hat{\Sigma}^{-1}\hat{U}^T.$$

(A6): Rank One Updates: A can be written as a sum of rank one matrices.

Proof. By direct computation, if A has $r \le \min\{m, n\}$ nonzero singular values, then

$$A = \hat{U}\hat{\Sigma}\hat{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots \sigma_r u_r v_r^T.$$

(A7): Compression: This is actually just an application of rank one updates. If we have a full rank matrix A, since $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0$, the leading columns of U and V with the leading singular values capture the majority of the energy in A. Therefore, we can store a low rank approximation of A by simply storing the first k columns of U and V and the first k singular values. Typically, the low rank approximation is very accurate.

12.4.1. *Computing the SVD.* In order to actually compute the SVD, we need to find the eigenpairs (λ_i, v_i) of $A^T A$. We will study two algorithms, power iteration and the QR algorithm to do this. However, we don't actually want to find the eigenpairs of $A^T A$ directly since $\text{cond}(A^T A) = (\text{cond}(A))^2$. Instead, we make a seemingly strange move of doubling the size of the matrix, but this new matrix B will have a condition number satsifyig $\text{cond}(B) = 2\text{cond}(A)$. Define $B \in \mathbb{R}^{(m+n)\times(m+n)}$ as the symmetric block matrix

$$B = \left[\begin{array}{c|c} 0 & A^T \\ \hline A & 0 \end{array}\right].$$

Let α be an eigenvalue of B with eigenvector $\begin{bmatrix} v \\ w \end{bmatrix}$ so that

$$\alpha \begin{bmatrix} v \\ w \end{bmatrix} = B \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} A^T w \\ Av \end{bmatrix}.$$

Notice that $A^T w = \alpha v$ and $Av = \alpha w$. Then,

$$A^T Av = A^T(\alpha w) = \alpha A^T w = \alpha(\alpha v) = \alpha^2 v.$$

So v is an eigenvector of $A^T A$ with eigenvalue $\lambda = \alpha^2$.

When the size A makes it feasible, we want to run the QR algorithm on B to find the eigenpairs (λ_i, v_i).