Theorem (The Spectral Theorem)

Let A be a real symmetric matrix. Then

1. *The eigenvalues of A are real.*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
3. *There exists a real diagonal matrix D and an orthogonal matrix Q such that*

$$\boxed{A = QDQ^T}$$

The matrix A^TA is special. It is a symmetric matrix, so the Spectral Theorem applies – it's eigenvalues are real and its eigenvectors are orthogonal. It is also used to find what is called the singular value decomposition, which gives us the best bases for the four fundamental subspaces.

Lemma

$A^T A$ and A have the same nullspaces.

Lemma

If A has LI columns, $A^T A$ is positive definite. Generally, $A^T A$ is nonnegative definite.

Now we are ready to prove the singular value decomposition theorem.

Theorem

Let A be an $m \times n$ matrix with real entries. Then there exists an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ diagonal matrix Σ such that

$$\boxed{A = U\Sigma V^T} \tag{$\star$}$$

where

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

The numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, arranged in decreasing order, are called the singular values of A. The decomposition $(\star)$ is called the singular value decomposition (SVD) of A.

$$\boxed{A\mathbf{u}_i = \sigma_i \mathbf{v}_i}$$

Proof.

Since A^TA is symmetric nonnegative definite, it has r positive eigenvalues and $n-r$ zero eigenvalues. Let $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$ be the positive eigenvalues of A^TA, arranged in decreasing order. Let $v_1, \ldots, v_r$ be the orthonormal eigenvectors for $\sigma_1^2, \ldots, \sigma_r^2$, and let $v_{r+1}, \ldots, v_n$ be the orthonormal eigenvectors for the zero eigenvalue. So

$$A^TAv_i = \sigma_i^2 v_i, \quad i = 1, \ldots r, \qquad A^TAv_i = 0, \quad i = r+1, \ldots, n$$

Now we *define* the vectors u_i as follows:

$$u_i = \frac{1}{\sigma_i} Av_i, \quad i = 1, \ldots, r$$

Notice that

$$(u_i, u_i) = \frac{1}{\sigma_i^2}(Av_i, Av_i) = \frac{1}{\sigma_i^2}(v_i, A^TAv_i) = \frac{1}{\sigma_i^2}(v_i, \sigma_i^2 v_i) = \frac{\sigma_i^2}{\sigma_i^2}(v_i, v_i) = 1$$

So the u_i's are necessarily length one. Moreover, by the same reasoning,

$$(u_i, u_j) = \frac{\sigma_j^2}{\sigma_i \sigma_j}(v_i, v_j) = 0$$

if $i \neq j$. Thus, $u_1, \ldots, u_j$ are orthonormal. Complete this to an orthonormal basis of $\mathbb{R}^m$: $\{u_1, \ldots, u_m\}$. Now let V be the matrix with columns v_i and let U be the matrix with columns u_i. Since $Av_i = \sigma_i u_i$, we have

$$AV = U\Sigma$$

$\square$

NOTES:

1. The singular values σ_i are the square roots of the eigenvalues of $A^T A$. They are all positive.
2. The vectors v_i are called the *right singular vectors*, and are the eigenvectors of $A^T A$. The u_i are the *left singular vectors*, and are the eigenvectors of AA^T.
3. If A is complex, the same result applies, but U and V are unitary:

   $$\boxed{A = U\Sigma V^*}$$

   for unitary U and V. The singular values are always positive.
4. It is useful to write the SVD as a sum of rank one matrices:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{0.1}$$

Example

Find the SVD of $A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$.

The SVD has myriad applications, from genomics to image processing. One of the uses of the SVD is to find something we can call the inverse for *any* matrix! Given the SVD $A = U\Sigma V^T$, we define the pseudo-inverse of Σ to be

$$\Sigma^+ = \begin{pmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

Then the pseudo-inverse of any matrix is given by

$$\boxed{A^+ = V\Sigma^+ U^T} \tag{0.2}$$

Or, usefully, the pseudo-inverse can be written as the sum of rank one matrices:

$$A^+ = \sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^T$$

Example

Show that the least squares solution to $Ax = b$ of minimal norm is

$$x = A^+ b$$

and the projection onto the range of A is given by

$$P = AA^+$$

Another reason for the importance of the SVD is that it gives us a way to approximate a matrix with a low rank matrix. This is the basis for many image compression algorithms.

Theorem

Let A be an $m \times n$ matrix. Then the closest rank k matrix to A is the matrix given by taking the first k singular values and setting the rest to zero. That is, let

$$A_k = U \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} V^T \tag{0.3}$$

Then A_k is the closest rank k matrix to A. If B is any other $m \times n$ rank k matrix then

$$\|A - B\| \geq \|A - A_k\|$$