1 Symmetric Matrices

A matrix A is called "symmetric" is $A^T = A$. If A is $m \times n$ then A^T is $n \times m$ so symmetric matrices must be square.

Symmetric matrices enjoy two useful properties:

1. They are diagonalizable (i.e., you can always find an eigenbasis for a symmetric matrix A)

2. Eigenspaces for different eigenvalues are orthogonal to one another.

Property 1 takes a while to prove. Property 2 has a simple proof that I covered in class but isn't something you need to know for this course.

For example,

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

is symmetric. Its eigenvalues are $\phi = \frac{1+\sqrt{5}}{2}$ and $-1/\phi = \frac{1-\sqrt{5}}{2}$. An eigenbasis is

$$\begin{pmatrix} \phi \\ 1 \end{pmatrix}, \begin{pmatrix} -1/\phi \\ 1 \end{pmatrix}$$

and these two eigenvectors are orthogonal, as expected:

$$\begin{pmatrix} \phi \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1/\phi \\ 1 \end{pmatrix} = 0$$

For any matrices A and B, note that $(AB)^T = B^T A^T$. To see this observe that the jith entry of AB is the jth row of A times the ith row of B, so the ijth entry of $(AB)^T$ is the jth row of A times the ith column of B, which is the jth column of A^T times the ith row of B^T, which is the ijth entry of $B^T A^T$. Of course, an easier way to "see this proof" is to exchange the rows and columns of the entire matrix multiplication all at once:

$$\begin{array}{cc} & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 2 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array} \longrightarrow \begin{array}{cc} & \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 2 & 0 \\ 3 & 1 \\ 1 & 1 \end{pmatrix} \end{array}$$

In particular, for any matrix A (maybe not even square), then $(A^T A)^T = A^T (A^T)^T = A^T A$, so $A^T A$ is always symmetric. Moreover its eigenvalues are always nonnegative: if $A^T A v = \lambda v$ then on the one hand

$$A^T A v \cdot v = \lambda v \cdot v = \lambda \|v\|^2$$

but on the other hand

$$A^T A v \cdot v = (A^T A v)^T v = v^T A^T A v = (Av)^T A v = \|Av\|^2$$

so

$$\lambda = \frac{\|Av\|^2}{\|v\|^2}$$

is a square, hence nonnegative.

2 Singular Vectors and Singular Values

Suppose A is an $n \times n$ matrix that is diagonalizable. Then you can find a basis $v_1, \dots, v_n$ such that A multiplies each v_i by a scalar λ_i.

Not all matrices are diagonalizable. However, the following remarkable fact is true for all matrices[1]. Given an $m \times n$ matrix A you can find an orthonormal basis $v_1, \dots, v_n$ of $\mathbb{R}^n$, an orthonormal basis $u_1, \dots, u_m$ of $\mathbb{R}^m$, and numbers $\sigma_1, \dots, \sigma_{\min(n,m)}$ such that $Av_i = \sigma_i u_i$. That is, you can find orthonormal bases of the domain and codomain so that A takes the vectors in the former to multiples of the vectors in the latter.

The vectors $v_1, \dots, v_n$ and the vectors $u_1, \dots, u_m$ are called singular vectors of A and the numbers $\sigma_1, \dots, \sigma_{\min(n,m)}$ are called the singular values of A.

To compute the singular vectors and singular values of A, follow the following recipe:

- Find an orthonormal eigenbasis of $A^T A$. This is $(v_1, \dots, v_n)$. Note that this is possible because $A^T A$ is symmetric.
- Reorder the basis so that $(v_1, \dots, v_k)$ are the eigenvectors corresponding to nonzero eigenvalues.
- Set $\sigma_i = \sqrt{\lambda_i}$. This is possible since the eigenvalues of $A^T A$ are nonnegative.
- For $1 \le i \le k$ set $u_i = \frac{1}{\sigma_i} A v_i$.
- Extend $(u_1, \dots, u_k)$ to an orthonormal basis $(u_1, \dots, u_m)$ of $\mathbb{R}^m$.

For example let

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Then

$$A^T A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The eigenvalues are 3, 1, and 0. You can easily find that an eigenbasis is

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

and these eigenvectors are orthogonal (as expected). To turn the eigenbasis into an orthonormal eigenbasis, normalize each vector:

$$v_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \ v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \ v_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

[1] This fact is proved at the end of these notes but the proof isn't something you need to know.

Set

$$u_1 = \frac{Av_1}{\sqrt{3}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$
$$u_2 = \frac{Av_2}{\sqrt{1}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

Since (u_1, u_2) forms a basis of $\mathbb{R}^2$, no more u_is are necessary. The singular values are $\sqrt{3}$ and 1.

As another example, let B be the transpose of the previous example:

$$B = \begin{pmatrix}1 & 0\\1 & 1\\0 & 1\end{pmatrix}$$

Then

$$B^T B = \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}$$

and the two eigenvalues of this matrix are 3 and 1. An orthonormal eigenbasis is

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Therefore set

$$u_1 = \frac{Bv_1}{\sqrt{3}} = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\2\\1\end{pmatrix}, \quad u_2 = \frac{Bv_2}{\sqrt{1}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}$$

(u_1, u_2) doesn't form a basis of $\mathbb{R}^3$, so add on u_3 to make orthonormal basis of $\mathbb{R}^3$. To compute u_3, note that an the orthogonal complement of $\text{span}(u_1, u_2)$ is the line given by the equations

$$\begin{cases} x + 2y + z = 0 \\ x - z = 0 \end{cases}$$

of which a basis is $(1, -1, 1)$. Therefore

$$u_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\end{pmatrix}$$

works. The singular vectors are (v_1, v_2) and (u_1, u_2, u_3). The singular values are $\sqrt{3}$ and 1.

Note that these singular vectors and values are familiar from the last example. Transposing a matrix doesn't change the singular values and exchanges the singular vectors in the domain and the codomain. This fact won't be proved in this class.

3 Singular Value Decomposition

The existence of singular vectors and singular values is important for the following reason: you can decompose any $m \times n$ matrix A in a particularly nice way:

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

where V and U are matrices whose columns are $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_m)$ (the singular vectors) and the diagonal entries of Σ are the singular values, i.e.,

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_m & 0 & \cdots & 0 \end{pmatrix} \text{ or } \Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

To see that this works, note that $V^T = V^{-1}$ since the columns of V are orthonormal and hence $V^T V$ is the identity matrix. Therefore an equivalent decomposition is $A = U\Sigma V^{-1}$. Then V^{-1} takes v_i to e_i, Σ scales e_i by σ_i, and U takes $\sigma_i e_i$ to $\sigma_i u_i$. Since a matrix is defined by what it does on a basis and A acts on the basis $(v_1, \ldots, v_n)$ in the exact same way, then $A = U\Sigma V^{-1}$.

Here's example that illustrates the utility of this decomposition. Let

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

Then

$$V = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$\sigma_1 = \sqrt{10}, \ \sigma_2 = 0$$

so that

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

In decimal approximations:

$$\sigma_1 = 3.1623$$

$$\sigma_2 = 0$$

$$v_1 = \begin{pmatrix} 0.8944 \\ 0.4472 \end{pmatrix}$$

$$v_2 = \begin{pmatrix} 0.4472 \\ -.8944 \end{pmatrix}$$

$$u_1 = \begin{pmatrix} 0.7071 \\ -.7071 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 0.7071 & 0.7071 \end{pmatrix}.$$

Consider changing the matrix for A just a little bit

$$B := \begin{pmatrix} 2.1 & 1.3 \\ -2.2 & -1.1 \end{pmatrix}$$

Then the singular values and singular vectors of B can be computed:

$$\sigma_1 = 3.4821$$

$$\sigma_2 = 0.1580$$

$$v_1 = \begin{pmatrix} 0.8732 \\ 0.4874 \end{pmatrix}$$

$$v_2 = \begin{pmatrix} 0.4874 \\ -.8732 \end{pmatrix}$$

$$u_1 = \begin{pmatrix} 0.7056 \\ .7086 \end{pmatrix}$$

$$u_2 = \begin{pmatrix} 0.7086 \\ 0.7056 \end{pmatrix}.$$

As might be expected, the singular values and singular vectors didn't change by much. What's interesting here is that, since σ_2 is so small,

$$\Sigma = \begin{pmatrix} 3.4821 & 0 \\ 0 & 0.1580 \end{pmatrix}$$

can be approximated by

$$\widetilde{\Sigma} = \begin{pmatrix} 3.4821 & 0 \\ 0 & 0 \end{pmatrix}$$

This is a simpler map, algebraically, since its image is one-dimensional. If $B = U\Sigma V^T$, define $\widetilde{B} = U\widetilde{\Sigma}V^T$. If B represents some observed data or relationships, then $\widetilde{B}$ represents a lower-dimensional approximation of the same data or relationships. Here's an example where that happens.

The singular values and singular vectors are used in image processing. Consider, for example, an 8×6 pixel image of the number 0. It might be represented in a matrix in the following form:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Each entry, viewed as a number from 0 to 1, represents how gray the pixel is, where 0 is completely white and 1 is completely black. Viewed as a transformation $\mathbb{R}^6 \to \mathbb{R}^8$, the image of the matrix is dimension 2, and therefore you'd expect two nonzero singular values.

If the image of the number 0 had some noise, the matrix might look like

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .98 & 1 & .98 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ .06 & 1 & 0 & .02 & 1 & 0 \\ 0 & 1 & .01 & 0 & 1 & 0 \\ 0 & 1 & .01 & 0 & .99 & 0 \\ 0 & 1 & .96 & .98 & 1 & 0 \\ 0 & 0 & 0 & 0 & .05 & 0 \end{pmatrix}$$

The dimension of the image of this matrix is not 2. You'd expect many nonzero singular values. In fact

$$\sigma_1 = 3.68$$

$$\sigma_2 = 1.4974$$

$$\sigma_3 = 0.0549$$

$$\sigma_4 = 0.0376$$

$$\sigma_5 = 0.0191$$

$$\sigma_6 = 0.$$

(The vectors $(v_1, \ldots, v_6)$ and $(u_1, \ldots, u_8)$ can also be computed, though I won't list them here.) While this new matrix has a 5-dimensional image, its image is still "approximately" 2-dimensional: two of singular values are much larger than the others. What this means is that this matrix can be approximated by the transformation

$$U\widetilde{\Sigma}V^T$$

where

$$\widetilde{\Sigma} = \begin{pmatrix} 3.68 & & & & & \\ & 1.4974 & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{pmatrix}$$

(and the rest of the entries equal to 0). This transformation can be simply summarized as:

$$v_1 \mapsto 3.68u_1$$
$$v_2 \mapsto 1.4974u_2$$
$$v_3 \mapsto 0$$
$$v_4 \mapsto 0$$
$$v_5 \mapsto 0$$
$$v_6 \mapsto 0$$

One would naively expect $U\widetilde{\Sigma}V^T$ to resemble the old matrix, and indeed, here it is:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .98 & .99 & .99 & 1 & 0 \\ .02 & 1 & 0 & 0 & 1 & 0 \\ .01 & 1 & .01 & .01 & 1 & 0 \\ .01 & 1 & .01 & .01 & 1 & 0 \\ .01 & 1 & .01 & .01 & .99 & \\ 0 & 1 & .97 & .97 & 1 & 0 \\ 0 & .02 & 0 & 0 & .02 & 0 \end{pmatrix}.$$

The '0' is still easy to see, but this matrix can be encoded in fewer pieces of information. One need only specify the vectors v_1, v_2, u_1, and u_2 and the two singular values s_1 and s_2. This is a total of $6+6+8+8+1+1 = 30$ pieces of information, less than the 48 required to represent the original matrix. As the matrices become much larger, this sort of approximation becomes more and more efficient.

Another place where SVD is useful is in statistics. As seen yesterday, it is helpful to encode the observations of p variables over N samples (e.g., 2 doses of medicine administered to each of 100 mice) in a $N \times p$ matrix X. The columns of X span a (typically) p-dimensional subspace of $\mathbb{R}^N$. A very small singular value of X means there's an element which is almost in the kernel of X. This means that there's almost a linear relation among the columns of X, and so the p-dimensional subspace of $\mathbb{R}^N$ can be approximated by a smaller dimensional subspace. Suppose that $p - k$ of the singular values are "small". Exactly what "small" is up for you to decide. Then you can write

$$X = U\Sigma V^T$$

and let $\overline{\Sigma}$ be the columns of Σ corresponding to the k largest singular values. If k is much less than p, then solving your problem with the $N \times k$ matrix $U\overline{\Sigma}$ instead of X is more computationally tractable. It is almost the same problem because the span of the columns of $U\overline{\Sigma}$ is almost the same as the span of the columns of X.

4 Proof of the Singular Value Decomposition

Let $(v_1, \ldots, v_n)$ be an orthonormal eigenbasis of A^TA. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues. Since A^TA is symmetric, such an eigenbasis always exists. Moreover, A^TA is positive semidefinite:

$$A^TAv \cdot v = \|Av\|^2 \geq 0$$

so that λ_i is real and nonnegative for all i.

Reorder the eigenbasis so that $v_1, \ldots, v_k$ are the eigenvectors with nonzero eigenvalues. For $1 \leq i \leq k$, define

$$u_i := \frac{Av_i}{\sqrt{\lambda_i}}.$$

Note that $\{Av_1, \ldots, Av_k\}$ are linearly independent: if

$$\sum_{i=1}^{k} a_i A v_i = 0$$

then

$$\sum_{i=1}^{k} a_i A^T A v_i = 0 \Rightarrow a_i \lambda_i v_i = 0$$
$$\Rightarrow a_i \lambda_i = 0 \ \forall i \Rightarrow a_i = 0 \ \forall i.$$

Since for $k + 1 \leq i \leq n$, $A^TAv_i = 0 \Rightarrow Av_i = 0$, the set $\{v_1, \ldots, v_k\}$ spans $\mathrm{im}(A)$. Hence $(A(v_1), \ldots, A(v_k))$ forms a basis for $\mathrm{im}(A)$. Since

$$Av_i \cdot Av_j = A^TAv_i \cdot v_j = \lambda_i v_i \cdot v_j$$

and $(v_1, \ldots, v_n)$ is orthonormal, then $(Av_1, \ldots, Av_k)$ is an orthogonal collection. For $1 \leq i \leq k$, let

$$u_i := \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sqrt{\lambda_i}}.$$

For $k+1 \leq i \leq m$, let $u_{k+1}, \ldots, u_m$ be an orthonormal extension to a basis of $\mathbb{R}^m$. Then

$$A(v) = A\left(\sum_{i=1}^{n}(v \cdot v_i)v_i\right) = \sum_{i=1}^{n}(v \cdot v_i)A(v_i) = \sum_{i=1}^{k}\sqrt{\lambda_i}(v \cdot v_i)u_i$$

$$= \sum_{i=1}^{\min(m,n)}\sqrt{\lambda_i}(v \cdot v_i)u_i$$

where the last equality uses the fact that $\sqrt{\lambda_i} = 0$ for $k+1 \leq i \leq m$.