7.2 Eigenvalues and eigenvectors of real symmetric matrices[1]

Result: in elementary linear algebra courses, it is shown that a real symmetric (so square) matrix $\mathbf{A}$ always has real eigenvalues and the eigenvectors of such a matrix may always be chosen to form an orthonormal set of size n.

Denote the eigenvectors of $\mathbf{A}$ by $\mathbf{u}_i$ with corresponding eigenvalue λ_i, so that

$$\mathbf{A}\mathbf{u}_i = \mathbf{u}_i\lambda_i \tag{3}$$

(note that I purposely write the right hand side this order. You can consider it as matrix multiplication of a $n \times 1$ matrix with a 1×1 matrix. Of course, the result is the same as standard scalar multiplication of a matrix $\lambda_i\mathbf{u}_i$.)

Write the column vectors $\mathbf{u}_i, \ldots, \mathbf{u}_n$ as the columns of a square matrix:

$$\mathbf{U} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

Now the relations described by Equation 3 can be written

$$\mathbf{AU} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \ldots & \mathbf{A}\mathbf{u}_n \\ \vdots & \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \lambda_1\mathbf{u}_1 & \lambda_1\mathbf{u}_2 & \ldots & \lambda_1\mathbf{u}_n \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \mathbf{UD}$$

where $\mathbf{D}$ is the diagonal matrix with eigenvalues on the diagonal, $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Since $\mathbf{U}$ is orthogonal, it is invertible with $\mathbf{U}^{-1} = \mathbf{U}^T$ so

$$\mathbf{A} = \mathbf{UDU}^T.$$

This can be rewritten as

$$\mathbf{A} = \sum_{k=1}^{n} \lambda_k\mathbf{u}_k\mathbf{u}_k^T,$$

a decomposition that we have seen before that is worth considering further. First, you can check that the decomposition is actually correct by showing that the multiplying the eigenvectors $\mathbf{u}_i$ by $\mathbf{A}$ and $\sum_{k=1}^{n} \lambda_k\mathbf{u}_k\mathbf{u}_k^T$ produces equivalent results.

Second, it means that the action of multiplying an arbitrary n-vector $\mathbf{x}$ by the real symmetric matrix $\mathbf{A}$, so $\mathbf{Ax} = (\sum_{k=1}^{n} \lambda_k\mathbf{u}_k\mathbf{u}_k^T)\mathbf{x} = \sum_{k=1}^{n} \mathbf{u}_k\lambda_k\mathbf{u}_k^T\mathbf{x}$ can be understood as comprising three steps:

[1] This section is taken, with few modifications, from notes written by Sze Tan for Physics 707: Inverse Problems.

1. It resolves the input vector along each of the eigenvectors $\mathbf{u}_k$, the component of the input vector along the kth eigenvector being given by $\mathbf{u}_k^T\mathbf{x}$,
2. The amount along the kth eigenvector is multiplied by the eigenvalue λ_k,
3. The product tells us how much of the kth eigenvector $\mathbf{u}_k$ is present in the product $\mathbf{Ax}$.

A schematic diagram of this process is shown in Figure 3.

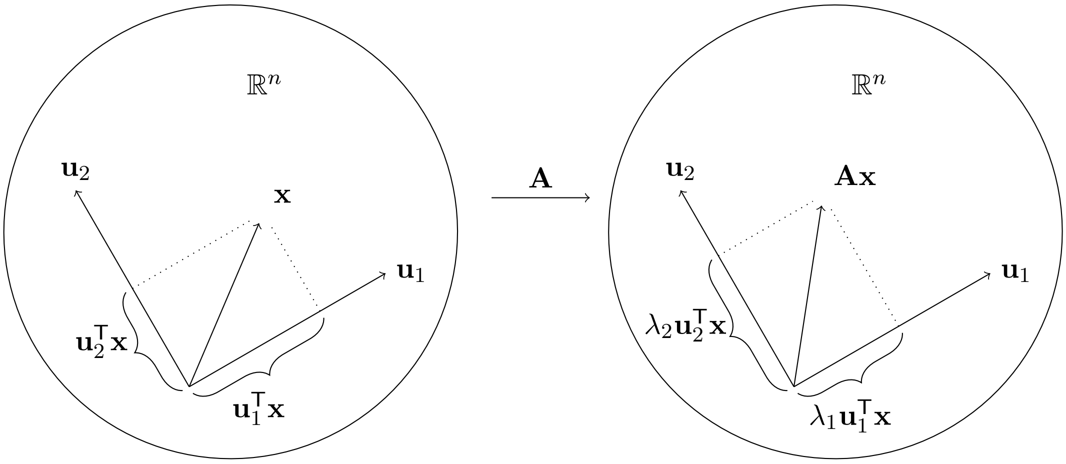

Figure 3: Effect of a real symmetric matrix $\mathbf{A}$ of size n on a vector $\mathbf{x}$. Only two of the orthogonal eigenvectors are shown.

Note that this is a special case of *diagonalisation*. A matrix $\mathbf{A}$ is diagonalisable if $\mathbf{A} = \mathbf{PDP}^{-1}$ (or, equivalently, $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$) for some diagonal matrix $\mathbf{D}$ and some matrix $\mathbf{P}$. In the case discussed above, where $\mathbf{A}$ is real and symmetric, $\mathbf{A}$ is diagonalisable with $\mathbf{P} = \mathbf{U}$ and $\mathbf{U}^{-1} = \mathbf{U}^\mathsf{T}$.

8 Singular Value Decomposition (SVD)

Singular Value Decomposition is a method of factorising any ordinary rectangular $m \times n$ matrix. It is most frequently applied to problems where $m \geq n$ (more equations than unknowns).

It has applications in signal processing, pattern recognition, and statistics where it is used for least squares data fitting, regularised inverse problems, finding pseudoinverses and performing principal component analysis (PCA). The application areas are many and varied but include computational tomography, seismology, weather forecast, image compression, image denoising, genetic analyses and more.

It is particularly useful when a given set of linear equations is singular or very close to singular in which case conventional solutions (e.g. by LU decomposition) are either not available or produce senseless results (due to the problems being ill-posed). In these cases, SVD can diagnose and, in some cases, solve the problem giving an useful numerical answer (though not necessarily the expected one!).

8.1 Overview of an SVD

SVD represents an ordinary $m \times n$ matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{UDV}^\mathsf{T}$ where:

- **U** : an $m \times m$ column-orthogonal matrix; its m columns are the m eigenvectors $\mathbf{u}$ of the $m \times m$ matrix $\mathbf{AA}^\mathsf{T}$. The vectors $\{\mathbf{u}\}$ are known as the *left singular vectors* of $\mathbf{A}$.
- **V** : an $n \times n$ orthogonal matrix; its n columns are the eigenvectors $\mathbf{v}$ of the $n \times n$ matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$. The vectors $\{\mathbf{v}\}$ are known as the *right singular vectors* of $\mathbf{A}$.
- **D** : an $m \times n$ matrix whose only non-zero elements are the first r entries on the diagonal where r is the rank of $\mathbf{A}$ and $d_{kk} = \sigma_k = \sqrt{\lambda_k} = \sqrt{\mu_k}$ where λ_k is the eigenvalue associated with $\mathbf{v}_k$ and μ_k is the eigenvalue associated with $\mathbf{u}_k$.

The *singular values*, σ_k are ordered so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$.
Since $\mathbf{A} = \mathbf{UDV}^\mathsf{T}$, we can write

$$\begin{aligned} \mathbf{A} &= \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^T \qquad (4) \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\mathsf{T} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\mathsf{T} + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\mathsf{T}. \end{aligned}$$

This representation suggests the approximation of $\mathbf{A}$ by the truncated series,

$$\widehat{\mathbf{A}}_\rho = \sum_{k=1}^{\rho} \sigma_k \mathbf{u}_k \mathbf{v}_k^T \text{ for } \rho < r.$$

Notice that when $m > n$ (that is, the problem is over-determined), there are at most n non-zero singular values. In this case, we can truncate the matrix $\mathbf{U}$ to be $m \times n$ and the matrix $\mathbf{D}$ to be a $n \times n$ diagonal matrix. This leaves the sum in Equation 4 unaltered as the rows or columns that are removed contribute nothing to that sum. In the following example, we employ this strategy.

Example: Find the SVD of the matrix $\mathbf{A}$ where

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

Solution: First find the eigenvectors and eigenvalues of $\mathbf{A}\mathbf{A}^\mathsf{T}$ and $\mathbf{A}^\mathsf{T}\mathbf{A}$. Since $\mathbf{A}$ is 3×2, we need only find the top two eigenvalues and eigenvectors of each of these matrices.

$$\mathbf{A}^\mathsf{T}\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

which has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$. The associated eigenvectors are, respectively,

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Notice that the eigenvectors have been normalised.
Similarly,

$$\mathbf{A}\mathbf{A}^\mathsf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

The top two eigenvalues are $\mu_1 = 3$ and $\mu_2 = 1$. Notice that these are the same as the top two eigenvalues of $\mathbf{A}^\mathsf{T}\mathbf{A}$. The associated eigenvectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

The singular values are given by $\sigma_i = \sqrt{\lambda_i}$, so $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$.
We can thus write $\mathbf{A}$ as

$$\begin{aligned} \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} &= \underbrace{[\mathbf{u}_1\ \mathbf{u}_2]}_{\mathbf{U}} \underbrace{\mathrm{diag}(\sigma_1, \sigma_2)}_{\mathbf{D}} \underbrace{\begin{bmatrix} \mathbf{v}_1^\mathsf{T} \\ \mathbf{v}_2^\mathsf{T} \end{bmatrix}}_{\mathbf{V}^\mathsf{T}} \\ &= \underbrace{\begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} \\ 2/\sqrt{6} & 0 \\ 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{D}} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{\mathbf{V}^\mathsf{T}}. \end{aligned}$$

The matrix approximation $\widehat{\mathbf{A}}_1$ is calculated as follows:

$$\begin{aligned} \widehat{\mathbf{A}}_1 &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\mathsf{T} \\ &= \sqrt{3} \cdot \tfrac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \cdot \tfrac{1}{\sqrt{2}} [1\ \ 1] = \tfrac{1}{2} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} [1\ \ 1] = \begin{bmatrix} 0.5 & 0.5 \\ 1 & 1 \\ 0.5 & 0.5 \end{bmatrix} \end{aligned}$$

while the approximation can be extended to $\widehat{\mathbf{A}}_2 (= \mathbf{A})$ by

$$\begin{aligned}
\widehat{\mathbf{A}}_2 &= \widehat{\mathbf{A}}_1 + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathsf{T}} \\
&\equiv \begin{bmatrix} 0.5 & 0.5 \\ 1 & 1 \\ 0.5 & 0.5 \end{bmatrix} + 1 \cdot \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tfrac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \end{bmatrix} \\
&\equiv \begin{bmatrix} 0.5 & 0.5 \\ 1 & 1 \\ 0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} -0.5 & 0.5 \\ 0 & 0 \\ 0.5 & -0.5 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}}_{\mathbf{A}}.
\end{aligned}$$