W07-05

Slide 01: In this unit, we will introduce the topic of eigenvalues and eigenvectors of a matrix.

Slide 02: To motivate the study of eigenvalues, consider the following real life example. Suppose it is known that the movement of people between the rural and urban district is as follows. Every year, 1% of the existing population in the rural district moves into the urban district while 4% of the existing population in the urban district moves in the opposite direction into the rural district.

(#)

In order to study this mathematically, we assume that the total population is a constant.

(#)

We would like to find out what will be the distribution of the total population in the long run. In other words, we would like to know how many percent of the population will be in each of the two districts. How can we study this mathematically?

Slide 03: Let us represent the urban population after n years by a_n. Similarly, let b_n be the rural population after n years.

(#)

For positive integers n, it is clear what a_n and b_n means. What about a_0 and b_0? We would let a_0 and b_0 represent the urban and rural population at the beginning of this study. You can also think of them as initial population in the two districts.

Slide 04: So how can we translate the information that we have about the population's movement every year into mathematical statements?

(#)

Before we do that, let us put both numbers, a_n and b_n into a 2×1 matrix, denoted by $\boldsymbol{x_n}$. The matrix would then contain the population distribution after n years.

(#)

In the same way we define $\boldsymbol{x_n}$, we can have $\boldsymbol{x_{n-1}}$ to contain the population distribution after $(n-1)$ years.

(#)

We are now ready to translate the movement information into equations involving our population distribution. So how is a_n dependent on a_{n-1} and b_{n-1}?

(#)

Well, we already know that 96% of the urban population continues to live in the urban district in the following year. We also know that 1% of the rural population will move into the urban district. Thus, we have $a_n = 0.96a_{n-1} + 0.01b_{n-1}$

(#)

Likewise, how can the rural population after n years be derived from the urban and rural population in year $(n-1)$?

(#)

Once again, it should be easy exercise to figure out that b_n is written as b_n equal $0.04a_{n-1} + 0.99b_{n-1}$.

(#)

The coefficients from both the equations can now be used to form a matrix.

Slide 05: Let $\boldsymbol{A}$ be the 2×2 matrix as shown. You can see that the entries in $\boldsymbol{A}$ are the coefficients from the two equations we have written down that relates $\boldsymbol{x_n}$ and $\boldsymbol{x_{n-1}}$.

(#)

With this choice of $\boldsymbol{A}$ we can now represent the two equations, which forms a linear system, as a matrix equation $\boldsymbol{x_n} = \boldsymbol{A}\boldsymbol{x_{n-1}}$. This is a useful equation as it relates the population distribution after n years with that after $n-1$ years. To get the population distribution matrix after n years, we simply premultiply the population distribution matrix after $n-1$ years by the square matrix $\boldsymbol{A}$.

(#)

If we can relate $\boldsymbol{x_n}$ and $\boldsymbol{x_{n-1}}$ in the way described above, we can certainly relate $\boldsymbol{x_{n-1}}$ and $\boldsymbol{x_{n-2}}$ in a similar manner.

(#)

Upon substituting $\boldsymbol{x_{n-1}} = \boldsymbol{A}\boldsymbol{x_{n-2}}$ into the first equation, we now see that $\boldsymbol{x_n}$ is equal to $\boldsymbol{A}^2$ premultiplied to $\boldsymbol{x_{n-2}}$.

(#)

We can continue to relate $\boldsymbol{x_{n-2}}$ and $\boldsymbol{x_{n-3}}$ as follows

(#)

and then substitute into the original equation to now obtain $\boldsymbol{x_n}$ equals to $\boldsymbol{A}^3$ premultiplied to $\boldsymbol{x_{n-3}}$. We are now seeing a pattern and as we continue in this manner,

(#)

we will arrive at the equation $\boldsymbol{x_n} = \boldsymbol{A}^n\boldsymbol{x_0}$. This is an important equation which tells us that if we know what is the initial population distribution $\boldsymbol{x_0}$, premultiplying $\boldsymbol{A}^n$ to $\boldsymbol{x_0}$ will allow us to compute the population distribution after n years, in the form of the matrix $\boldsymbol{x_n}$.

Slide 06: Now conceptually, we are now able to find out the population distribution after 1000 years, **provided** we know what is the matrix $\boldsymbol{A}^{1000}$.

Slide 07: Obviously, computing $\boldsymbol{A}^{1000}$ is tedious and almost impossible to do by hand. Suppose you have been told, and then verified that the matrix $\boldsymbol{A}$ can be written as the product of three matrices as shown. Notice that the first and third matrices are inverses of each other while the second matrix is a diagonal matrix.

(#)

Let us represent these three matrices as $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{P}^{-1}$. So now we have $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$.

(#)

The fact that $\boldsymbol{P}$ is invertible and $\boldsymbol{D}$ is diagonal is going to be crucial in helping us compute the powers of $\boldsymbol{A}$.

(#)

Now $\boldsymbol{A}^n$ would simply be n copies of $\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ multiplied together.

(#)

Writing down the n copies of $\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$,

(#)

and removing the brackets since matrix multiplication is associative,

(#)

first observe that now we have $\boldsymbol{P}^{-1}\boldsymbol{P}$ occuring in the matrix product. Remember that $\boldsymbol{P}^{-1}\boldsymbol{P}$ is equal to the identity matrix.

(#)

Thus the product of the matrices can now be simplified to $\boldsymbol{P}$ premultiplied to a total of n $\boldsymbol{D}$s and then with $\boldsymbol{P}^{-1}$ postmultiplied to the result.

(#)

The expression now simplifies to $\boldsymbol{P}\boldsymbol{D}^n\boldsymbol{P}^{-1}$.

(#)

Now some of you may wonder, how can this be any simpler than before since in order to compute $\boldsymbol{A}^{1000}$, we need to compute $\boldsymbol{D}^{1000}$. Is computing $\boldsymbol{D}^{1000}$ easier to do?

Slide 08: Let us examine how we can compute $\boldsymbol{D}^{1000}$.

(#)

It is indeed convenient due to the diagonal nature of $\boldsymbol{D}$ that, for example, $\boldsymbol{D}^2$ is simply computed by having the diagonal entries of the matrix squared, as shown. This is one of the properties of diagonal matrices that we are using to assist in the computation.

(#)

Likewise, $\boldsymbol{D}^3$ is obtained by raising the diagonal entries of $\boldsymbol{D}$ to the power of 3.

(#)

In general, we have the following expression for $\boldsymbol{D}^n$, notice that 1^n is 1 for all positive integers n.

(#)

Recall that we have obtained the expression for $\boldsymbol{A}^n$ as $\boldsymbol{P}\boldsymbol{D}^n\boldsymbol{P}^{-1}$ and now we can write down $\boldsymbol{D}^n$ explicitly as follows.

(#)

Now in the context of the population example, we would like to know what is the distribution in the long run, meaning that as n becomes very large. So we see that as $n \to \infty$, $(0.95)^n$ would approach 0 and therefore $\boldsymbol{A}^n$ would converge to the product of the three matrices as shown. $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ remains since it is independent of n and $\boldsymbol{D}^n$ converges to $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Slide 09: Now if we put in matrices $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$,

(#)

we will be able to evaluate the 2×2 matrix that $\boldsymbol{A}^n$ will converge to as $n \to \infty$. Note that there is still a little work to be done here in finding the inverse of $\boldsymbol{P}$. Of course you can do that using the Gaussian elimination method as described in an earlier unit.

(#)

Now that we know what $\boldsymbol{A}^n$ converges to as n becomes large, we will know what $\boldsymbol{x_n}$ will converge to as n becomes large, since $\boldsymbol{x_n} = \boldsymbol{A}^n\boldsymbol{x_0}$.

(#)

The answer is simply the 2×2 matrix we have obtained premultiplied to the initial population distribution matrix.

(#)

Now a_n is found to be $0.2(a_0 + b_0)$

(#)
while b_n is $0.8(a_0 + b_0)$.
(#)
In other words we can now conclude that after a long time, the urban population will be 20% of the total population while the other 80% will be staying in the rural district.
(#)
Now that we have come to the conclusion of this problem, can you identify which was the crucial step that enabled us to do so?

Slide 10: It was the part where we were able to write $\boldsymbol{A}$ as the product of three matrices $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{P}^{-1}$. Firstly, having $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ in the positions we have allowed us to simplify the expression when we have n copies of $\boldsymbol{PDP}^{-1}$ multiplied together. Secondly, the diagonal nature of $\boldsymbol{D}$ allowed us to compute the powers of $\boldsymbol{D}$ easily, unlike the matrix $\boldsymbol{A}$. We are now ready to formally define eigenvectors and eigenvalues of a matrix, which will eventually lead to the discussion of when a square matrix can be written in such a convenient manner like we had for the matrix $\boldsymbol{A}$ in this example.

Slide 11: Let $\boldsymbol{A}$ be a square matrix of order n. A **non zero** column vector $\boldsymbol{u}$ in $\mathbb{R}^n$ is called an eigenvector of $\boldsymbol{A}$ if $\boldsymbol{Au} = \lambda\boldsymbol{u}$ for some scalar λ.
(#)
What this means is that when $\boldsymbol{A}$ is premultiplied to $\boldsymbol{u}$, the effect is that $\boldsymbol{u}$ gets scaled by a factor of λ. Note that it is important to remember that $\boldsymbol{u}$ must be non zero in this definition.
(#)
The scalar λ also has a name. It is called an eigenvalue of $\boldsymbol{A}$ and we say that $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ **associated with** the eigenvalue λ.

Slide 12: Before we conclude this unit, take note of the following. If $\boldsymbol{u}$ is known to be an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ, what about vectors like $2\boldsymbol{u}$ or $-1.5\boldsymbol{u}$? or any scalar multiple of $\boldsymbol{u}$ in general?
(#)
Let us consider $2\boldsymbol{u}$ and premultiply $\boldsymbol{A}$ to it. Since 2 is a scalar, it can be taken outside the bracket.
(#)
We now apply what we know about $\boldsymbol{u}$, namely that it is an eigenvector of $\boldsymbol{A}$ associated with λ, so we write $\boldsymbol{Au}$ as $\lambda\boldsymbol{u}$.
(#)
We now bring 2 back into the bracket
(#)
and see that we have shown $\boldsymbol{A}(2\boldsymbol{u})$ is equal to $\lambda(2\boldsymbol{u})$. This means that $2\boldsymbol{u}$ is also an eigenvector of $\boldsymbol{A}$ associated with the same eigenvalue λ.
(#)
We can show similarly, that $(-1.5)\boldsymbol{u}$ and $300\boldsymbol{u}$ will also be eigenvectors of $\boldsymbol{A}$ associated with λ.
(#)

In fact, all scalar multiples of $\boldsymbol{u}$ will also be eigenvectors of $\boldsymbol{A}$ associated with the same eigenvalue λ. In the following units, we will attempt to find all the eigenvalues and eigenvectors of any given square matrix $\boldsymbol{A}$.

Slide 13: In summary,

(#)

we started with a real life example on population movement that led to the need to compute powers of a square matrix.

(#)

We defined what is an eigenvalue of a square matrix and also eigenvectors of the matrix associated with a particular eigenvalue.