bookboon LEARNING

Elementary Linear Algebra: Part II

Kenneth Kuttler

KENNETH KUTTLER

ELEMENTARY LINEAR ALGEBRA: PART II

Elementary Linear Algebra: Part II
3rd edition

ISBN 978-87-403-3128-8

CONTENTS

To see Chapter 1-9 download

Elementary Linear Algebra: Part I

10 A FEW FACTORIZATIONS

10.1 DEFINITION OF AN LU FACTORIZATION

An LU factorization of a matrix involves writing the given matrix as the product of a lower triangular matrix which has the main diagonal consisting entirely of ones L and an upper triangular matrix U in the indicated order. This is the version discussed here but it is sometimes the case that the L has numbers other than 1 down the main diagonal. It is still a useful concept. The L goes with "lower" and the U with "upper". It turns out many matrices can be written in this way and when this is possible, people get excited about slick ways of solving the system of equations, $A\mathbf{x} = \mathbf{y}$. It is for this reason that you want to study the LU factorization. It allows you to work only with triangular matrices. It turns out that it takes about half as many operations to obtain an LU factorization as it does to find the row reduced echelon form.

First it should be noted not all matrices have an LU factorization and so we will emphasize the techniques for achieving it rather than formal proofs.

Example 10.11 *Can you write* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *in the form LU as just described?*

To do so you would need

$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a & b \\ xa & xb+c \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore, $b = 1$ and $a = 0$. Also, from the bottom rows, $xa = 1$ which can't happen and have $a = 0$. Therefore, you can't write this matrix in the form LU. It has no LU factorization. This is what we mean above by saying the method lacks generality.

10.2 FINDING AN LU FACTORIZATION BY INSPECTION

Which matrices have an LU factorization? It turns out it is those whose row reduced echelon form can be achieved without switching rows and which only involve row operations of type 3 in which row j is replaced with a multiple of row i added to row j for $i < j$.

Example 10.2.1 *Find an LU factorization of* $A = \begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{pmatrix}$

One way to find the LU factorization is to simply look for it directly. You need

$$\begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & z & 1 \end{pmatrix} \begin{pmatrix} a & d & h & j \\ 0 & b & e & i \\ 0 & 0 & c & f \end{pmatrix}.$$

Then multiplying these you get

$$\begin{pmatrix} a & d & h & j \\ xa & xd+b & xh+e & xj+i \\ ya & yd+zb & yh+ze+c & yj+iz+f \end{pmatrix}$$

and so you can now tell what the various quantities equal. From the first column, you need $a=1, x=1, y=2$. Now go to the second column. You need $d=2, xd+b=3$ so $b=1, yd+zb=3$ so $z=-1$. From the third column, $h=0, e=2, c=6$. Now from the fourth column, $j=2, i=-1, f=-5$. Therefore, an LU factorization is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 6 & -5 \end{pmatrix}.$$

You can check whether you got it right by simply multiplying these two.

10.3 USING MULTIPLIERS TO FIND AN LU FACTORIZATION

There is also a convenient procedure for finding an LU factorization. It turns out that it is only necessary to keep track of the **multipliers** which are used to row reduce to upper triangular form. This procedure is described in the following examples.

Example 103.1 *Find an LU factorization for* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}$

Write the matrix next to the identity matrix as shown.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}.$$

The process involves doing row operations to the matrix on the right while simultaneously updating successive columns of the matrix on the left. First take -2 times the first row and add to the second in the matrix on the right.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 1 & 5 & 2 \end{pmatrix}$$

Note the way we updated the matrix on the left. We put a 2 in the second entry of the first column because we used -2 times the first row added to the second row. Now replace the third row in the matrix on the right by -1 times the first row added to the third. Notice that the product of the two matrices is unchanged and equals the original matrix. This is because a row operation was done on the original matrix to get the matrix on the right and then on the left, it was multiplied by an elementary matrix which "undid" the row operation which was done.

The next step is

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 3 & -1 \end{pmatrix}$$

Again, the product is unchanged because we just did and then undid a row operation. Finally, we will add the second row to the bottom row and make the following changes

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 0 & -11 \end{pmatrix}.$$

At this point, we stop because the matrix on the right is upper triangular. An LU factorization is the above.

The justification for this gimmick will be given later in a more general context.

Example 10.3.2 *Find an* LU *factorization for* $A = \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 3 & 1 & 3 & 2 \\ 1 & 0 & 1 & 1 & 2 \end{pmatrix}.$

►►

We will use the same procedure as above. However, this time we will do everything for one column at a time. First multiply the first row by (-1) and then add to the last row. Next take (-2) times the first and add to the second and then (-2) times the first and add to the third.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & -2 & 0 & -1 & 1 \end{pmatrix}.$$

This finishes the first column of L and the first column of U. As in the above, what happened was this. Lots of row operations were done and then these were undone by multiplying by the matrix on the left. Thus the above product equals the original matrix. Now take $-(1/4)$ times the second row in the matrix on the right and add to the third followed by $-(1/2)$ times the second added to the last.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 1/4 & 1 & 0 \\ 1 & 1/2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & 0 & -1 & -1/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 3/2 \end{pmatrix}$$

This finishes the second column of L as well as the second column of U. Since the matrix on the right is upper triangular, stop. The LU factorization has now been obtained. This technique is called Dolittle's method.

This process is entirely typical of the general case. The matrix U is just the first upper triangular matrix you come to in your quest for the row reduced echelon form using only the row operation which involves replacing a row by itself added to a multiple of another row. The matrix L is what you get by updating the identity matrix as illustrated above.

You should note that for a square matrix, the number of row operations necessary to reduce LU to form is about half the number needed to place the matrix in row reduced echelon form. This is why an LU factorization is of interest in solving systems of equations.

10.4 SOLVING SYSTEMS USING AN LU FACTORIZATION

One reason people care about the LU factorization is it allows the quick solution of systems of equations. Here is an example.

Example 10.4.1. *Suppose you want to find the solutions to*

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Of course one way is to write the augmented matrix and grind away. However, this involves more row operations than the computation of the LU factorization and it turns out that the LU factorization can give the solution quickly. Here is how. The following is an LU factorization for the matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

Let $U\mathbf{x} = \mathbf{y}$ and consider $L\mathbf{y} = \mathbf{b}$ where in this case, $\mathbf{b} = (1,2,3)^T$. Thus

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

which yields very quickly that $\mathbf{y} = \begin{pmatrix} 1 & -2 & 2 \end{pmatrix}^T$. Now you can find $\mathbf{x}$ by solving $U\mathbf{x} = \mathbf{y}$. Thus in this case,

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

which yields

$$\mathbf{x} = \begin{pmatrix} \frac{7}{5}t - \frac{3}{5} & \frac{9}{5} - \frac{11}{5}t & t & -1 \end{pmatrix}^T, t \in \mathbb{R}.$$

10.5 JUSTIFICATION FOR THE MULTIPLIER METHOD

Why does the multiplier method work for finding the LU factorization? Suppose A is a matrix which has the property that the row reduced echelon form for A may be achieved using only the row operations which involve replacing a row with itself added to a multiple of another row. It is not ever necessary to switch rows. Thus every row which is replaced using this row operation in obtaining the echelon form may be modified by using a row which is above it.

Lemma 10.5.1 *Let L be a lower (upper) triangular matrix $m \times m$ which has ones down the main diagonal. Then L^{-1} also is a lower (upper) triangular matrix which has ones down the main diagonal. In the case that L is of the form*

$$L = \begin{pmatrix} 1 & & & \\ a_1 & 1 & & \\ \vdots & & \ddots & \\ a_n & & & 1 \end{pmatrix} \tag{10.1}$$

where all entries are zero except for the left column and main diagonal, it is also the case that L^{-1} is obtained from L by simply multiplying each entry below the main diagonal in L with . -1 The same is true if the single nonzero column is in another position.

Proof: Consider the usual setup for finding the inverse $\begin{pmatrix} L & I \end{pmatrix}$. Then each row operation done to L to reduce to row reduced echelon form results in changing only the entries in I below the main diagonal. In the special case of L given in 10.1 or the single nonzero column is in another position, multiplication by -1 as described in the lemma clearly results in L^{-1}. ■

For a simple illustration of the last claim,

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & a & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -a & 1 \end{pmatrix}$$

Now let A be an $m \times n$ matrix, say

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

and assume can be row reduced to an upper triangular form using only row operation 3. Thus, in particular, $a_{11} \neq 0$. Multiply on the left by $E_1 =$

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{m1}}{a_{11}} & 0 & \cdots & 1 \end{pmatrix}$$

This is the product of elementary matrices which make modifications in the first column only. It is equivalent to taking $-a_{21}/a_{11}$ times the first row and adding to the second. Then taking $-a_{31}/a_{11}$ times the first row and adding to the third and so forth. The quotients in the first column of the above matrix are the multipliers. Thus the result is of the form

$$E_1 A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \cdots & a'_{mn} \end{pmatrix}$$

By assumption, $a'_{22} \neq 0$ and so it is possible to use this entry to zero out all the entries below it in the matrix on the right by multiplication by a matrix of the form $E_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & E \end{pmatrix}$ where E is an $(m-1) \times (m-1)$ matrix of the form

$$E = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{a'_{32}}{a'_{22}} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a'_{m2}}{a'_{22}} & 0 & \cdots & 1 \end{pmatrix}$$

Again, the entries in the first column below the 1 are the multipliers. Continuing this way, zeroing out the entries below the diagonal entries, finally leads to

$$E_{m-1}E_{n-2}\cdots E_1A = U$$

where U is upper triangular. Each E_j has all ones down the main diagonal and is lower triangular. Now multiply both sides by the inverses of the E_j in the reverse order This yields

$$A = E_1^{-1}E_2^{-1}\cdots E_{m-1}^{-1}U$$

By Lemma 10.5.1, this implies that the product of those E_j^{-1} is a lower triangular matrix having all ones down the main diagonal.

The above discussion and lemma gives the justification for the multiplier method. The expressions

$$-a_{21}/a_{11}, -a_{31}/a_{11}, \cdots - a_{m1}/a_{11}$$

denoted respectively by $M_{21}, \cdots, M_{m1}$ to save notation which were obtained in building E_1 are the multipliers. Then according to the lemma, to find E_1^{-1} you simply write

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{pmatrix}$$

Similar considerations apply to the other E_j^{-1}. Thus L is a product of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -M_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -M_{m1} & 0 & \cdots & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 & & \cdots & 0 \\ 0 & 1 & & \cdots & 0 \\ \vdots & 0 & & \ddots & \vdots \\ 0 & \cdots & & -M_{m(m-1)} & 1 \end{pmatrix}$$

each factor having at most one nonzero column, the position of which moves from left to right in scanning the above product of matrices from left to right. It follows from Theorem 8.1.9 about the effect of multiplying on the left by an elementary matrix that the above product is of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -M_{21} & 1 & \cdots & 0 & 0 \\ \vdots & -M_{32} & \ddots & \vdots & \vdots \\ -M_{(M-1)1} & \vdots & \cdots & 1 & 0 \\ -M_{M1} & -M_{M2} & \cdots & -M_{MM-1} & 1 \end{pmatrix}$$

In words, beginning at the left column and moving toward the right, you simply insert, into the corresponding position in the identity matrix, -1 times the multiplier which was used to zero out an entry in that position below the main diagonal in A, while retaining the main diagonal which consists entirely of ones. This is L.

10.6 THE PLU FACTORIZATION

As indicated above, some matrices don't have an LU factorization. Here is an example.

$$M = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} \tag{10.2}$$

In this case, there is another factorization which is useful called a PLU factorization. Here P is a permutation matrix.

Example 10.6.1. *Find a* PLU *factorization for the above matrix in 10.2.*

Proceed as before trying to find the row echelon form of the matrix. First add -1 times the first row to the second row and then add -4 times the first to the third. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 \\ 0 & -5 & -11 & -7 \end{pmatrix}$$

There is no way to do only row operations involving replacing a row with itself added to a multiple of another row to the matrix on the right in such a way as to obtain an upper triangular matrix. Therefore, consider the original matrix with the bottom two rows switched.

$$M' = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} = PM$$

Now try again with this matrix. First take -1 times the first row and add to the bottom row and then take -4 times the first row and add to the second row. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

The matrix on the right is upper triangular and so the LU factorization of the matrix M' has been obtained above.

Thus $M' = PM = LU$ where and are given above. Notice that $P^2 = I$ and therefore, $M = P^2M = PLU$ and so

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

This process can always be followed and so there always exists a PLU factorization of a given matrix even though there isn't always an LU factorization.

Example 10.6.2 *Use the* PLU *factorization of* $M \equiv \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix}$ *to solve the system* $M\mathbf{x} = \mathbf{b}$ *where* $\mathbf{b} = (1, 2, 3)^T$.

Let $U\mathbf{x} = \mathbf{y}$ and consider $PL\mathbf{y} = \mathbf{b}$. In other words, solve,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Multiplying both sides by P gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

and so

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Now $U\mathbf{x} = \mathbf{y}$ and so it only remains to solve

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

which yields

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^T =$$
$$= \begin{pmatrix} \frac{7}{5}t + \frac{1}{5} & \frac{9}{10} - \frac{11}{5}t & t & -\frac{1}{2} \end{pmatrix}^T : t \in \mathbb{R}.$$

10.7 THE QR FACTORIZATION

As pointed out above, the LU factorization is not a mathematically respectable thing because it does not always exist. There is another factorization which does always exist. Much more can be said about it than I will say here. At this time, I will only deal with real matrices and so the inner product will be the usual real dot product. Letting A be an $m \times n$ real matrix and letting $(\cdot,\cdot)$ denote the usual real inner product,

$$\begin{aligned} (A\mathbf{x},\mathbf{y}) &= \sum_i (A\mathbf{x})_i y_i = \sum_i \sum_j A_{ij} x_j y_i = \sum_j \sum_i \left(A^T\right)_{ji} y_i x_j \\ &= \sum_j \left(A^T \mathbf{y}\right)_j x_j = \left(\mathbf{x}, A^T \mathbf{y}\right) \end{aligned}$$

Thus, when you take the matrix across the comma, you replace with a transpose.

Definition 10.7.1 *An $n \times n$ real matrix Q is called an orthogonal matrix if*

$$QQ^T = Q^T Q = I.$$

Thus an orthogonal matrix is one whose inverse is equal to its transpose.

From the above observation,

$$|Q\mathbf{x}|^2 = (Q\mathbf{x}, Q\mathbf{x}) = \left(\mathbf{x}, Q^T Q\mathbf{x}\right) = (\mathbf{x}, I\mathbf{x}) = (\mathbf{x}, \mathbf{x}) = |\mathbf{x}|^2$$

This shows that orthogonal transformations preserve distances. Conversely you can also show that if you have a matrix which does preserve distances, then it must be orthogonal.

Example 10.7.2 *One of the most important examples of an orthogonal matrix is the so called Householder matrix. You have* $\mathbf{v}$ *a unit vector and you form the matrix*

$$I - 2\mathbf{v}\mathbf{v}^T$$

This is an orthogonal matrix which is also symmetric. To see this, you use the rules of matrix operations.

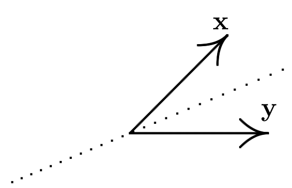
$$\begin{aligned} \left(I - 2\mathbf{v}\mathbf{v}^T\right)^T &= I^T - \left(2\mathbf{v}\mathbf{v}^T\right)^T \\ &= I - 2\mathbf{v}\mathbf{v}^T \end{aligned}$$

so it is symmetric. Now to show it is orthogonal,

$$\begin{aligned} \left(I - 2\mathbf{v}\mathbf{v}^T\right)\left(I - 2\mathbf{v}\mathbf{v}^T\right) &= I - 2\mathbf{v}\mathbf{v}^T - 2\mathbf{v}\mathbf{v}^T + 4\mathbf{v}\mathbf{v}^T\mathbf{v}\mathbf{v}^T \\ &= I - 4\mathbf{v}\mathbf{v}^T + 4\mathbf{v}\mathbf{v}^T = I \end{aligned}$$

because $\mathbf{v}^T\mathbf{v} = \mathbf{v} \cdot \mathbf{v} = |\mathbf{v}|^2 = 1$. *Therefore, this is an example of an orthogonal matrix.*

Next consider the problem illustrated in the following picture.



Find an orthogonal matrix Q which switches the two vectors taking $\mathbf{x}$ to $\mathbf{y}$ and $\mathbf{y}$ to $\mathbf{x}$.

Procedure 10.7.3 *Given two vectors* $\mathbf{x}, \mathbf{y}$ *such that* $|\mathbf{x}| = |\mathbf{y}| \neq 0$ *but* $\mathbf{x} \neq \mathbf{y}$ *and you want an orthogonal matrix* Q *such that* $Q\mathbf{x} = \mathbf{y}$ *and* $Q\mathbf{y} = \mathbf{x}$. *The thing which works is the Householder matrix*

$$Q \equiv I - 2\frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|^2}(\mathbf{x} - \mathbf{y})^T$$

Here is why this works.

$$\begin{aligned} Q(\mathbf{x}-\mathbf{y}) &= (\mathbf{x}-\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y}) \\ &= (\mathbf{x}-\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}|\mathbf{x}-\mathbf{y}|^2 = \mathbf{y}-\mathbf{x} \end{aligned}$$

$$\begin{aligned} Q(\mathbf{x}+\mathbf{y}) &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}+\mathbf{y}) \\ &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}((\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})) \\ &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}\left(|\mathbf{x}|^2 - |\mathbf{y}|^2\right) = \mathbf{x}+\mathbf{y} \end{aligned}$$

Hence

$$\begin{aligned} Q\mathbf{x}+Q\mathbf{y} &= \mathbf{x}+\mathbf{y} \\ Q\mathbf{x}-Q\mathbf{y} &= \mathbf{y}-\mathbf{x} \end{aligned}$$

Adding these equations, $2Q\mathbf{x} = 2\mathbf{y}$ and subtracting them yields $2Q\mathbf{y} = 2\mathbf{x}$.

Definition 10.7.4 Let A be an $m \times n$ matrix. Then a QR factorization of A consists of two matrices, Q orthogonal and R upper triangular or in other words equal to zero below the main diagonal such that $A = QR$.

With the solution to this simple problem, here is how to obtain a QR factorization for any matrix A Let

$$A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$$

where the $\mathbf{a}_i$ are the columns. If $\mathbf{a}_1 = \mathbf{0}$, let $Q_1 = I$. If $\mathbf{a}_1 \neq \mathbf{0}$, let

$$\mathbf{b} \equiv \begin{pmatrix} |\mathbf{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and form the Householder matrix

$$Q_1 \equiv I - 2\frac{(\mathbf{a}_1 - \mathbf{b})}{|\mathbf{a}_1 - \mathbf{b}|^2} (\mathbf{a}_1 - \mathbf{b})^T$$

As in the above problem $Q_1 \mathbf{a}_1 = \mathbf{b}$ and so

$$Q_1 A = \begin{pmatrix} |\mathbf{a}_1| & * \\ \mathbf{0} & A_2 \end{pmatrix}$$

where A_2 is a $m - 1 \times n - 1$ matrix. Now find in the same way as was just done a $n - 1 \times n - 1$ matrix $\widehat{Q}_2$ such that

$$\widehat{Q}_2 A_2 = \begin{pmatrix} * & * \\ \mathbf{0} & A_3 \end{pmatrix}$$

Let

$$Q_2 \equiv \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix}.$$

Then

$$Q_2 Q_1 A = \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix} \begin{pmatrix} |\mathbf{a}_1| & * \\ \mathbf{0} & A_2 \end{pmatrix}$$
$$= \begin{pmatrix} |\mathbf{a}_1| & * & * \\ \vdots & * & * \\ 0 & \mathbf{0} & A_3 \end{pmatrix}$$

Continuing this way until the result is upper triangular, you get a sequence of orthogonal matrices $Q_p Q_{p-1} \cdots Q_1$ such that

$$Q_p Q_{p-1} \cdots Q_1 A = R \tag{10.3}$$

where R is upper triangular.

Now if Q_1 and Q_2 are orthogonal, then from properties of matrix multiplication,

$$Q_1 Q_2 (Q_1 Q_2)^T = Q_1 Q_2 Q_2^T Q_1^T = Q_1 I Q_1^T = I$$

and similarly

$$(Q_1 Q_2)^T Q_1 Q_2 = I.$$

Thus the product of orthogonal matrices is orthogonal. Also the transpose of an orthogonal matrix is orthogonal directly from the definition. Therefore, from 10.3

$$A = (Q_p Q_{p-1} \cdots Q_1)^T R \equiv QR,$$

where Q is orthogonal. This suggests the proof of the following theorem.

Theorem 10.7.5 *Let A be any real $m \times n$ matrix. Then there exists an orthogonal matrix Q and an upper triangular matrix R having nonnegative entries down the main diagonal such that*

$$A = QR$$

and this factorization can be accomplished in a systematic manner.

Proof: The theorem is clearly true if A is a $1 \times m$ matrix. Suppose it is true for A an $n \times m$ matrix. Thus, if A is any $n \times m$ matrix, there exists an orthogonal matrix Q such that

$$QA = R$$

where R is upper triangular. Suppose A is an $(n+1)\times m$ matrix. Then, as indicated above, there exists an orthogonal $(n+1)\times(n+1)$ matrix Q_1 such that

$$Q_1 A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{0} & A_1 \end{pmatrix}$$

where A_1 is $n \times m-1$ or else, in case $m=1$, the right side is of the form

$$\begin{pmatrix} a \\ \mathbf{0} \end{pmatrix}$$

and in this case, the conclusion of the theorem follows. If $m-1 \geq 1$, then by induction, there exists Q_2 an orthogonal $n \times n$ matrix such that $Q_2 A_1 = R_1$. Then

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix} Q_1 A = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix}\begin{pmatrix} a & \mathbf{b} \\ \mathbf{0} & A_1 \end{pmatrix} = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{0} & R_1 \end{pmatrix} \equiv R$$

Since $\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_2 \end{pmatrix} Q_1$ is orthogonal, being the product of two orthogonal matrices, the conclusion follows. ■

► ►

10.8 MATLAB AND FACTORIZATIONS

MATLAB can find an LU factorizaton of a matrix. Here is an example.

```
>> A=[1,3,2,4;-5,7,2,3;2,3,7,11;1,2,3,4]; [L,U,P]=lu(sym(A))
```

This will give a lower triangular matrix L with ones down the diagonal, an upper triangular matrix U, and a permutation matrix P such that $PA = LU$. Of course if the matrix A has an LU factorization, you will have $P = I$. This was the case here. After you have typed the above, you press enter and you get the following listed in a column. If you just type lu(A) , then the answer will come out in decimals.

L=	U=	P=
[1, 0, 0, 0]	[1, 3, 2, 4]	1 0 0 0
[-5, 1, 0, 0]	[0, 22, 12, 23]	0 1 0 0
[2, -3/22, 1, 0]	[0, 0, 51/11, 135/22]	0 0 1 0
[1, -1/22, 1/3, 1]	[0, 0, 0, -1]	0 0 0 1

MATLAB can also find the much more interesting QR factorization. Here is how to do it with an example.

```
>> A=[1,2,3;4,2,1;2,6,7;1,-4,2];[Q,R]=qr(A)
```

Then press enter and you get the following.

```
Q=                                      R=
-0.2132 0.1756 -0.2593 0.9255           -4.6904 -3.8376 -4.9036
-0.8528 -0.1892 0.4862 -0.0244          0 6.7285 3.4453
-0.4264 0.6485 -0.5139 -0.3653          0 0 -5.2043
-0.2132 -0.7161 -0.6575 -0.0974         0 0 0
```

If you want to see something horrible, replace qr(A) with qr(sym(A)). This way it gives the exact values. You can check your work by >>Q*Q' and press enter. The Q' means the conjugate transpose. Since everything is real here, this is just the transpose.

10.9 EXERCISES

1. Find an LU factorization of $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$

2. Find an LU factorization of $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 1 \\ 5 & 0 & 1 & 3 \end{pmatrix}$

3. Find an LU factorization of the matrix $\begin{pmatrix} 1 & -2 & -5 & 0 \\ -2 & 5 & 11 & 3 \\ 3 & -6 & -15 & 1 \end{pmatrix}$

4. Find an LU factorization of the matrix $\begin{pmatrix} 1 & -1 & -3 & -1 \\ -1 & 2 & 4 & 3 \\ 2 & -3 & -7 & -3 \end{pmatrix}$

5. Find an LU factorization of the matrix $\begin{pmatrix} 1 & -3 & -4 & -3 \\ -3 & 10 & 10 & 10 \\ 1 & -6 & 2 & -5 \end{pmatrix}$

6. Find an LU factorization of the matrix $\begin{pmatrix} 1 & 3 & 1 & -1 \\ 3 & 10 & 8 & -1 \\ 2 & 5 & -3 & -3 \end{pmatrix}$

7. Find an LU factorization of the matrix $\begin{pmatrix} 3 & -2 & 1 \\ 9 & -8 & 6 \\ -6 & 2 & 2 \\ 3 & 2 & -7 \end{pmatrix}$

8. Find an LU factorization of the matrix $\begin{pmatrix} -3 & -1 & 3 \\ 9 & 9 & -12 \\ 3 & 19 & -16 \\ 12 & 40 & -26 \end{pmatrix}$

9. Find an LU factorization of the matrix $\begin{pmatrix} -1 & -3 & -1 \\ 1 & 3 & 0 \\ 3 & 9 & 0 \\ 4 & 12 & 16 \end{pmatrix}$

10. Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y &= 5 \\ 2x + 3y &= 6 \end{aligned}$$

11. Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y + z &= 1 \\ y + 3z &= 2 \\ 2x + 3y &= 6 \end{aligned}$$

12. Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y + 3z &= 5 \\ 2x + 3y + z &= 6 \\ x - y + z &= 2 \end{aligned}$$

13. Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y + 3z &= 5 \\ 2x + 3y + z &= 6 \\ 3x + 5y + 4z &= 11 \end{aligned}$$

14. Is there only one LU factorization for a given matrix? **Hint:** Consider the equation

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

Look for all possible LU factorizations.

15. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \end{pmatrix}$.

16. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 4 & 2 & 4 & 1 \\ 1 & 2 & 1 & 3 & 2 \end{pmatrix}$.

17. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 4 & 1 \\ 3 & 2 & 1 \end{pmatrix}$.

18. Find a PLU factorization of $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 0 & 2 \\ 2 & 2 & 1 \end{pmatrix}$ and use it to solve the systems

a. $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 0 & 2 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$ b. $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \\ 1 & 0 & 2 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$

19. Find a PLU factorization of $\begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 1 & -2 & 0 \\ 2 & 3 & -1 & 2 \end{pmatrix}$ and use it to solve the systems

a. $\begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 1 & -2 & 0 \\ 2 & 3 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ b. $\begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 1 & -2 & 0 \\ 2 & 3 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$

20. Find a QR factorization for the matrix

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & -2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$$

21. Find a QR factorization for the matrix

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

22. If you had a QR factorization, $A = QR$, describe how you could use it to solve the equation $A\mathbf{x} = \mathbf{b}$. This is not usually the way people solve this equation. However, the QR factorization is of great importance in certain other problems, especially in finding eigenvalues and eigenvectors.
23. In this problem, is another explanation of the LU factorization. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 1 & 3 & 1 \end{pmatrix}$$

Review how to take the inverse of an elementary matrix. Then we can do the following.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & -2 \\ 1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 1 & 3 & 1 \end{pmatrix}$$

Next

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 1 & 3 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 1 & -2 \end{pmatrix} \end{aligned}$$

Next

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 1 & -2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -\frac{1}{4} & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & 0 & -4 \end{pmatrix} \end{aligned}$$

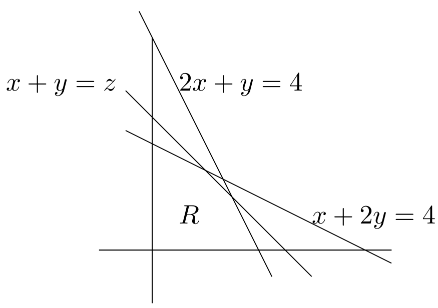
Using this example, describe why, when a matrix can be reduced to echelon form using only row operation 3, then it has an LU factorization.

11 LINEAR PROGRAMMING

11.1 SIMPLE GEOMETRIC CONSIDERATIONS

One of the most important uses of row operations is in solving linear program problems which involve maximizing a linear function subject to inequality constraints determined from linear equations. Here is an example. A certain hamburger store has 9000 hamburger patties to use in one week and a limitless supply of special sauce, lettuce, tomatoes, onions, and buns. They sell two types of hamburgers, the big stack and the basic burger. It has also been determined that the employees cannot prepare more than 9000 of either type in one week. The big stack, popular with the teenagers from the local high school, involves two patties, lots of delicious sauce, condiments galore, and a divider between the two patties. The basic burger, very popular with children, involves only one patty and some pickles and ketchup. Demand for the basic burger is twice what it is for the big stack. What is the maximum number of hamburgers which could be sold in one week given the above limitations?

Let x be the number of basic burgers and y the number of big stacks which could be sold in a week. Thus it is desired to maximize $z = x + y$ subject to the above constraints. The total number of patties is 9000 and so the number of patty used is $x + 2y$. This number must satisfy $x + 2y \leq 9000$ because there are only 9000 patty available. Because of the limitation on the number the employees can prepare and the demand, it follows $2x + y \leq 9000$. You never sell a negative number of hamburgers and so $x, y \geq 0$. In simpler terms the problem reduces to maximizing $z = x + y$ subject to the two constraints, $x + 2y \leq 9000$ and $2x + y \leq 9000$. This problem is pretty easy to solve geometrically. Consider the following picture in which labels the region described by the above inequalities and the line $z = x + y$ is shown for a particular value of z.



As you make z larger this line moves away from the origin, always having the same slope and the desired solution would consist of a point in the region, R which makes z as large as possible or equivalently one for which the line is as far as possible from the origin. Clearly this point is the point of intersection of the two lines, $(3000, 3000)$ and so the maximum value of the given function is 6000. Of course this type of procedure is fine for a situation in which there are only two variables but what about a similar problem in which there are very many variables. In reality, this hamburger store makes many more types of burgers than those two and there are many considerations other than demand and available patty. Each will likely give you a constraint which must be considered in order to solve a more realistic problem and the end result will likely be a problem in many dimensions, probably many more than three so your ability to draw a picture will get you nowhere for such a problem. Another method is needed. This method is the topic of this section. I will illustrate with this particular problem. Let $x_1 = x$ and $y = x_2$. Also let x_3 and x_4 be nonnegative variables such that

$$x_1 + 2x_2 + x_3 = 9000, \; 2x_1 + x_2 + x_4 = 9000.$$

To say that x_3 and x_4 are nonnegative is the same as saying $x_1 + 2x_2 \leq 9000$ and $2x_1 + x_2 \leq 9000$ and these variables are called slack variables at this point. They are called this because they "take up the slack". I will discuss these more later. First a general situation is considered.

11.2 THE SIMPLEX TABLEAU

Here is some notation.

Definition 11.2.1 *Let* $\mathbf{x}, \mathbf{y}$ *be vectors in* $\mathbb{R}^q$. *Then* $\mathbf{x} \leq \mathbf{y}$ *means for each* $i, x_i \leq y_i$.

The problem is as follows:

Let A be an $m \times (m+n)$ real matrix of rank m It is desired to find $\mathbf{x} \in \mathbb{R}^{n+m}$ such that $\mathbf{x}$ satisfies the constraints,

$$\mathbf{x} \geq \mathbf{0}, A\mathbf{x} = \mathbf{b} \tag{11.1}$$

and out of all such $\mathbf{x}$,

$$z \equiv \sum_{i=1}^{m+n} c_i x_i$$

is as large (or small) as possible. This is usually referred to as maximizing or minimizing z subject to the above constraints. First I will consider the constraints.

Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{n+m} \end{pmatrix}$. First you find a vector $\mathbf{x}^0 \geq \mathbf{0}$, $A\mathbf{x}^0 = \mathbf{b}$ such that n of the components of this vector equal 0. Letting $i_1, \cdots, i_n$ be the positions of $\mathbf{x}^0$ for which $x_i^0 = 0$, suppose also that $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$ is linearly independent for j_i the other positions of $\mathbf{x}^0$. Geometrically, this means that $\mathbf{x}^0$ is a corner of the feasible region, those $\mathbf{x}$ which satisfy the constraints. This is called a basic feasible solution. Also define

$$\begin{aligned} \mathbf{c}_B &\equiv (c_{j_1}, \cdots, c_{j_m}), \ \mathbf{c}_F \equiv (c_{i_1}, \cdots, c_{i_n}) \\ \mathbf{x}_B &\equiv (x_{j_1}, \cdots, x_{j_m}), \ \mathbf{x}_F \equiv (x_{i_1}, \cdots, x_{i_n}). \end{aligned}$$

and

$$z^0 \equiv z(\mathbf{x}^0) = \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \mathbf{c}_B \mathbf{x}_B^0$$

since $\mathbf{x}_F^0 = \mathbf{0}$. The variables which are the components of the vector $\mathbf{x}_B$ are called the **basic variables** and the variables which are the entries of $\mathbf{x}_F$ are called the **free variables**. You set $\mathbf{x}_F = \mathbf{0}$. Now $(\mathbf{x}^0, z^0)^T$ is a solution to

$$\begin{pmatrix} A & \mathbf{0} \\ -\mathbf{c} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}$$

along with the constraints $\mathbf{x} \geq \mathbf{0}$. Writing the above in augmented matrix form yields

$$\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix} \tag{11.2}$$

Permute the columns and variables on the left if necessary to write the above in the form

$$\begin{pmatrix} B & F & \mathbf{0} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_F \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \tag{11.3}$$

or equivalently in the augmented matrix form keeping track of the variables on the bottom as

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \\ \mathbf{x}_B & \mathbf{x}_F & 0 & 0 \end{pmatrix}. \tag{11.4}$$

Here B pertains to the variables $x_{i_1}, \cdots, x_{j_m}$ and is an $m \times m$ matrix with linearly independent columns, $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$, and F is an $m \times n$ matrix. Now it is assumed that

$$\begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{0} \end{pmatrix} = B\mathbf{x}_B^0 = \mathbf{b}$$

and since B is assumed to have rank m, it follows

$$\mathbf{x}_B^0 = B^{-1}\mathbf{b} \geq \mathbf{0}. \tag{11.5}$$

This is very important to observe. $B^{-1}\mathbf{b} \geq \mathbf{0}$! This is by the assumption that $\mathbf{x}^0 \geq \mathbf{0}$.

Do row operations on the top part of the matrix

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \tag{11.6}$$

and obtain its row reduced echelon form. Then after these row operations the above becomes

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix}. \tag{11.7}$$

where $B^{-1}\mathbf{b} \geq 0$. Next do another row operation in order to get a $\mathbf{0}$ where you see a $-\mathbf{c}_B$. Thus

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{pmatrix} \tag{11.8}$$

$$= \begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B \mathbf{x}_B^0 \end{pmatrix}$$

$$= \begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & z^0 \end{pmatrix} \tag{11.9}$$

The reason there is a z^0 on the bottom right corner is that $\mathbf{x}_F = 0$ and $\left(\mathbf{x}_B^0, \mathbf{x}_F^0, z^0\right)^T$ is a solution of the system of equations represented by the above augmented matrix because it is a solution to the system of equations corresponding to the system of equations represented by 11.6 and row operations leave solution sets unchanged. Note how attractive this is. The z_0 is the value of z at the point $\mathbf{x}^0$. The augmented matrix of 11.9 is called the simplex tableau and it is the beginning point for the simplex algorithm to be described a little later. It is very convenient to express the simplex tableau in the above form in which the variables are possibly permuted in order to have $\begin{pmatrix} I \\ \mathbf{0} \end{pmatrix}$ on the left side. However, as far as the simplex algorithm is concerned it is not necessary to be permuting the variables in this manner. Starting with 11.9 you could permute the variables and columns to obtain an augmented matrix in which the variables are in their original order. What is really required for the simplex tableau?

It is an augmented $m+1 \times m+n+2$ matrix which represents a system of equations which has the same set of solutions, $(\mathbf{x},z)^T$ as the system whose augmented matrix is

$$\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix}$$

(Possibly the variables for $\mathbf{x}$ are taken in another order.) There are linearly independent columns in the first $m+n$ columns for which there is only one nonzero entry, a 1 in one of the first m rows, the "simple columns", the other first $m+n$ columns being the "nonsimple columns". As in the above, the variables corresponding to the simple columns are $\mathbf{x}_B$ the basic variables and those corresponding to the nonsimple columns are $\mathbf{x}_F$, the free variables. Also, the top m entries of the last column on the right are nonnegative. This is the description of a simplex tableau.

In a simplex tableau it is easy to spot a basic feasible solution. You can see one quickly by setting the variables, $\mathbf{x}_F$ corresponding to the nonsimple columns equal to zero. Then the other variables, corresponding to the simple columns are each equal to a nonnegative entry in the far right column. Lets call this an **"obvious basic feasible solution"**. If a solution is obtained by setting the variables corresponding to the nonsimple columns equal to zero and the variables corresponding to the simple columns equal to zero this will be referred to as an **"obvious"** solution. Lets also call the first $m+n$ entries in the bottom row the "bottom left row". In a simplex tableau, the entry in the bottom right corner gives the value of the variable being maximized or minimized when the obvious basic feasible solution is chosen.

The following is a special case of the general theory presented above and shows how such a special case can be fit into the above framework. The following example is rather typical of the sorts of problems considered. It involves inequality constraints instead of $A\mathbf{x} = \mathbf{b}$. This is handled by adding in "slack variables" as explained below.

The idea is to obtain an augmented matrix for the constraints such that obvious solutions are also feasible. Then there is an algorithm, to be presented later, which takes you from one obvious feasible solution to another until you obtain the maximum.

Example 11.2.2 *Consider $z = x_1 - x_2$ subject to the constraints, $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2$, and $2x_1 + x_2 \leq 6, x_i \geq 0$. Find a simplex tableau for a problem of the form $\mathbf{x} \geq \mathbf{0}$, $A\mathbf{x} = \mathbf{b}$ which is equivalent to the above problem.*

You add in slack variables. These are positive variables, one for each of the first three constraints, which change the first three inequalities into equations. Thus the first three inequalities become $x_1 + 2x_2 + x_3 = 10, x_1 + 2x_2 - x_4 = 2,$ and $2x_1 + x_2 + x_5 = 6, x_1, x_2, x_3, x_4, x_5 \geq 0$

. Now it is necessary to find a basic feasible solution. You mainly need to find a positive solution to the equations,

$$\begin{array}{c} x_1 + 2x_2 + x_3 = 10 \\ x_1 + 2x_2 - x_4 = 2 \\ 2x_1 + x_2 + x_5 = 6 \end{array} .$$

the solution set for the above system is given by

$$x_2 = \frac{2}{3}x_4 - \frac{2}{3} + \frac{1}{3}x_5, x_1 = -\frac{1}{3}x_4 + \frac{10}{3} - \frac{2}{3}x_5, x_3 = -x_4 + 8.$$

An easy way to get a basic feasible solution is to let $x_4 = 8$ and $x_5 = 1$. Then a feasible solution is

$$(x_1, x_2, x_3, x_4, x_5) = (0, 5, 0, 8, 1).$$

It follows $z^0 = -5$ and the matrix 11.2, $\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix}$ with the variables kept track of on the bottom is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 & 1 & 0 & 6 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 \\ x_1 & x_2 & x_3 & x_4 & x_5 & 0 & 0 \end{pmatrix}$$

and the first thing to do is to permute the columns so that the list of variables on the bottom will have x_1 and x_3 at the end.

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 & 10 \\ 2 & -1 & 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 & 0 & 6 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ x_2 & x_4 & x_5 & x_1 & x_3 & 0 & 0 \end{pmatrix}$$

Next, as described above, take the row reduced echelon form of the top three lines of the above matrix. This yields

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \end{pmatrix}.$$

Now do row operations to

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}$$

to finally obtain

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

and this is a simplex tableau. The variables are $x_2, x_4, x_5, x_1, x_3, z$.

It isn't as hard as it may appear from the above. Lets not permute the variables and simply find an acceptable simplex tableau as described above.

Example 11.2.3 *Consider* $z = x_1 - x_2$ *subject to the constraints,* $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2$, *and* $2x_1 + x_2 \leq 6, x_i \geq 0$. *Find a simplex tableau.*

Adding in slack variables, an augmented matrix which is descriptive of the constraints is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 6 \\ 2 & 1 & 0 & 0 & 1 & 6 \end{pmatrix}$$

The obvious solution is not feasible because of that -1 in the fourth column. When you let $x_1, x_2 = 0$, you end up having $x_4 = -6$ which is negative. Consider the second column and select the 2 as a pivot to zero out that which is above and below the 2.

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 3 \end{pmatrix}$$

This one is good. When you let $x_1 = x_4 = 0$, you find that $x_2 = 3, x_3 = 4, x_5 = 3$. The obvious solution is now feasible. You can now assemble the simplex tableau. The first step is to include a column and row for z. This yields

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Now you need to get zeros in the right places so the simple columns will be preserved as simple columns in this larger matrix. This means you need to zero out the 1 in the third column on the bottom. A simplex tableau is now

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 0 & -1 & 0 & 1 & -4 \end{pmatrix}.$$

Note it is not the same one obtained earlier. There is no reason a simplex tableau should be unique. In fact, it follows from the above general description that you have one for each basic feasible point of the region determined by the constraints.

11.3 THE SIMPLEX ALGORITHM

11.3.1 MAXIMUMS

The simplex algorithm takes you from one basic feasible solution to another while maximizing or minimizing the function you are trying to maximize or minimize. Algebraically, it takes you from one simplex tableau to another in which the lower right corner either increases in the case of maximization or decreases in the case of minimization.

I will continue writing the simplex tableau in such a way that the simple columns having only one entry nonzero are on the left. As explained above, this amounts to permuting the variables. I will do this because it is possible to describe what is going on without onerous notation. However, in the examples, I won't worry so much about it. Thus, from a basic feasible solution, a simplex tableau of the following form has been obtained in which the columns for the basic variables, are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{11.10}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $(\mathbf{x}^0, z^0)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $(\mathbf{x}^0, z^0)$ is a basic feasible solution.

If $c_i < 0$ for some i, and if $F_{ji} \leq 0$ so that a whole column of $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ is ≤ 0 with the bottom entry < 0, then letting x_i be the variable corresponding to that column, you could leave all the other entries of $\mathbf{x}_F$ equal to zero but change to x_i be positive. Let the new vector be denoted by $\mathbf{x}'_F$ and letting $\mathbf{x}'_B = \mathbf{b} - F\mathbf{x}'_F$ it follows

$$\begin{aligned} (\mathbf{x}'_B)_k &= b_k - \sum_j F_{kj} (\mathbf{x}_F)_j \\ &= b_k - F_{ki} x_i \geq 0 \end{aligned}$$

Now this shows $(\mathbf{x}'_B, \mathbf{x}'_F)$ is feasible whenever $x_i > 0$ and so you could let x_i become arbitrarily large and positive and conclude there is no maximum for z because

$$z = (-c_i)\, x_i + z^0 \tag{11.11}$$

If this happens in a simplex tableau, you can say there is no maximum and stop.

What if $\mathbf{c} \geq \mathbf{0}$? Then $z = z^0 - \mathbf{c}\mathbf{x}_F$ and to satisfy the constraints, you need $\mathbf{x}_F \geq 0$. Therefore, in this case, z^0 is the largest possible value of z and so the maximum has been found. You stop when this occurs. Next I explain what to do if neither of the above stopping conditions hold.

The only case which remains is that some $c_i < 0$ and some $F_{ji} > 0$. You pick a column in $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ in which $c_i < 0$, usually the one for which c_i is the largest in absolute value. You pick $F_{ji} > 0$ as a pivot element, divide the j^{th} row by F_{ji} and then use to obtain zeros above F_{ji} and below F_{ji}, thus obtaining a new simple column. This row operation also makes exactly one of the other simple columns into a nonsimple column. (In terms of variables, it is said that a free variable becomes a basic variable and a basic variable becomes a free variable.) Now permuting the columns and variables, yields

$$\begin{pmatrix} I & F' & \mathbf{0} & \mathbf{b}' \\ \mathbf{0} & \mathbf{c}' & 1 & z^{0\prime} \end{pmatrix}$$

where $z^{0\prime} \geq z^0$ because $z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}}\right)$ and $c_i < 0$. If $\mathbf{b}' \geq 0$, you are in the same position you were at the beginning but now is larger. Now here is the **important** thing. You don't pick just any when you do these row operations. **You pick the positive one for which the row operation results in** $\mathbf{b}' \geq 0$. Otherwise the obvious basic feasible solution obtained by letting $\mathbf{x}'_F = \mathbf{0}$ will fail to satisfy the constraint that $\mathbf{x} \geq \mathbf{0}$.

How is this done? You need

$$b_k' \equiv b_k - \frac{F_{ki} b_j}{F_{ji}} \geq 0 \tag{11.12}$$

for each $k = 1, \cdots, m$ or equivalently,

$$b_k \geq \frac{F_{ki} b_j}{F_{ji}}. \tag{11.13}$$

Now if $F_{ki} \leq 0$ the above holds. Therefore, you only need to check F_{pi} for $F_{pi} > 0$. The pivot, F_{ji} is the one which makes the quotients of the form

$$\frac{b_p}{F_{pi}}$$

for all positive F_{pi} the smallest. This will work because for $F_{ki} > 0$,

Having gotten a new simplex tableau, you do the same thing to it which was just done and continue. As long as $\mathbf{b} > \mathbf{0}$, so you don't encounter the degenerate case, the values for z associated with setting $\mathbf{x}_F = \mathbf{0}$ keep getting strictly larger every time the process is repeated. You keep going until you find $\mathbf{c} \geq \mathbf{0}$. Then you stop. You are at a maximum. Problems can occur in the process in the so called degenerate case when at some stage of the process some $b_j = 0$. In this case you can cycle through different values for $\mathbf{x}$ with no improvement in z . This case will not be discussed here.

Example 11.3.1 *Maximize* $2x_1 + 3x_2$ *subject to the constraints* $x_1 + x_2 \geq 1, 2x_1 + x_2 \leq 6, x_1 + 2x_2 \leq 6$*,* $x_1, x_2 \geq 0$.

The constraints are of the form

$$\begin{aligned} x_1 + x_2 - x_3 &= 1 \\ 2x_1 + x_2 + x_4 &= 6 \\ x_1 + 2x_2 + x_5 &= 6 \end{aligned}$$

where the x_3, x_4, x_5 are the slack variables. An augmented matrix for these equations is of the form

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 1 & 0 & 6 \\ 1 & 2 & 0 & 0 & 1 & 6 \end{pmatrix}$$

Obviously the obvious solution is not feasible. It results in $x_3 < 0$. We need to exchange basic variables. Lets just try something.

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 5 \end{pmatrix}$$

Now this one is all right because the obvious solution is feasible. Letting $x_2 = x_3 = 0$, it follows that the obvious solution is feasible. Now we add in the objective function as described above.

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 0 & 5 \\ -2 & -3 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Then do row operations to leave the simple columns the same. Then

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 1 & 0 & 0 & 4 \\ 0 & 1 & 1 & 0 & 1 & 0 & 5 \\ 0 & -1 & -2 & 0 & 0 & 1 & 2 \end{pmatrix}$$

Now there are negative numbers on the bottom row to the left of the 1. Lets pick the first. (It would be more sensible to pick the second.) The ratios to look at are $5/1, 1/1$ so pick for the pivot the 1 in the second column and first row. This will leave the right column above the lower right corner nonnegative. Thus the next tableau is

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 5 \\ -1 & 0 & 2 & 0 & 1 & 0 & 4 \\ 1 & 0 & -3 & 0 & 0 & 1 & 3 \end{pmatrix}$$

There is still a negative number there to the left of the 1 in the bottom row. The new ratios are $4/2, 5/1$ so the new pivot is the 2 in the third column. Thus the next tableau is

$$\begin{pmatrix} \frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & 0 & 3 \\ \frac{3}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & 3 \\ -1 & 0 & 2 & 0 & 1 & 0 & 4 \\ -\frac{1}{2} & 0 & 0 & 0 & \frac{3}{2} & 1 & 9 \end{pmatrix}$$

Still, there is a negative number in the bottom row to the left of the 1 so the process does not stop yet. The ratios are $3/(3/2)$ and $3/(1/2)$ and so the new pivot is that $3/2$ in the first column. Thus the new tableau is

$$\begin{pmatrix} 0 & 1 & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & 2 \\ \frac{3}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & 3 \\ 0 & 0 & 2 & \frac{2}{3} & \frac{2}{3} & 0 & 6 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{4}{3} & 1 & 10 \end{pmatrix}$$

Now stop. The maximum value is 10. This is an easy enough problem to do geometrically and so you can easily verify that this is the right answer. It occurs when $x_4 = x_5 = 0, x_1 = 2, x_2 = 2, x_3 = 3$.

11.3.2 MINIMUMS

How does it differ if you are finding a minimum? From a basic feasible solution, a simplex tableau of the following form has been obtained in which the simple columns for the basic variables, $\mathbf{x}_B$ are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{11.14}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $(\mathbf{x}^0, z^0)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $(\mathbf{x}^0, z^0)$ is a basic feasible solution. So far, there is no change.

Suppose first that some $c_i > 0$ and $F_{ji} \leq 0$ for each j. Then let $\mathbf{x}'_F$ consist of changing x_i by making it positive but leaving the other entries of $\mathbf{x}_F$ equal to 0. Then from the bottom row,

$$z = -c_i x_i + z^0$$

and you let $\mathbf{x}'_B = \mathbf{b} - F\mathbf{x}'_F \geq \mathbf{0}$. Thus the constraints continue to hold when x_i is made increasingly positive and it follows from the above equation that there is no minimum for z. You stop when this happens.

Next suppose $\mathbf{c} \leq \mathbf{0}$. Then in this case, $z = z^0 - \mathbf{c}\mathbf{x}_F$ and from the constraints, $\mathbf{x}_F \geq 0$ and so and $-\mathbf{c}\mathbf{x}_F \geq 0$ so z^0 is the minimum value and you stop since this is what you are looking for.

What do you do in the case where some $c_i > 0$ and some $F_{ji} > 0$? In this case, you use the simplex algorithm as in the case of maximums to obtain a new simplex tableau in which $z^{0\prime}$ is smaller. You choose F_{ji} the same way to be the positive entry of the i^{th} column such that $b_p/F_{pi} \geq b_j/F_{ji}$ for all positive entries, F_{pi} and do the same row operations. Now this time,

$$z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}} \right) < z^0$$

As in the case of maximums no problem can occur and the process will converge unless you have the degenerate case in which some $b_j = 0$. As in the earlier case, this is most unfortunate when it occurs. You see what happens of course. z^0 does not change and the algorithm just delivers different values of the variables forever with no improvement.

To summarize the geometrical significance of the simplex algorithm, it takes you from one corner of the feasible region to another. You go in one direction to find the maximum and in another to find the minimum. For the maximum you try to get rid of negative entries of $\mathbf{c}$ and for minimums you try to eliminate positive entries $\mathbf{c}$, of where the method of elimination involves the auspicious use of an appropriate pivot element and row operations.

Now return to Example 11.2.2. It will be modified to be a maximization problem.

Example 11.3.2 *Maximize* $z = x_1 - x_2$ *subject to the constraints,*

$$x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2,$$

and $2x_1 + x_2 \leq 6, x_i \geq 0.$

Recall this is the same as maximizing $z = x_1 - x_2$ subject to

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & -1 & 0 \\ 2 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 6 \end{pmatrix}, \mathbf{x} \geq \mathbf{0},$$

the variables, x_3, x_4, x_5 being slack variables. Recall the simplex tableau was

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

with the variables ordered as x_2, x_4, x_5, x_1, x_3 and so $\mathbf{x}_B = (x_2, x_4, x_5)$ and

$$\mathbf{x}_F = (x_1, x_3).$$

Apply the simplex algorithm to the fourth column because $-\frac{3}{2} < 0$ and this is the most negative entry in the bottom row. The pivot is $3/2$ because $1/(3/2) = 2/3 < 5/(1/2)$. Dividing this row by $3/2$ and then using this to zero out the other elements in that column, the new simplex tableau is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{14}{3} \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & \frac{2}{3} & 1 & -\frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 & 0 & -1 & 1 & -4 \end{pmatrix}.$$

Now there is still a negative number in the bottom left row. Therefore, the process should be continued. This time the pivot is the $2/3$ in the top of the column. Dividing the top row by $2/3$ and then using this to zero out the entries below it,

$$\begin{pmatrix} \frac{3}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & 7 \\ -\frac{3}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & 3 \\ \frac{3}{2} & 0 & \frac{1}{2} & 0 & 0 & 1 & 3 \end{pmatrix}.$$

Now all the numbers on the bottom left row are nonnegative so the process stops. Now recall the variables and columns were ordered as x_2, x_4, x_5, x_1, x_3. The solution in terms of

x_1 and x_2 is $x_2 = 0$ and $x_1 = 3$ and $z = 3$. Note that in the above, I did not worry about permuting the columns to keep those which go with the basic variables on the left.

Here is a bucolic example.

Example 11.3.3 *Consider the following table.*

	F_1	F_2	F_3	F_4
iron	1	2	1	3
protein	5	3	2	1
folic acid	1	2	2	1
copper	2	1	1	1
calcium	1	1	1	1

This information is available to a pig farmer and denotes a particular feed. The numbers in the table contain the number of units of a particular nutrient contained in one pound of the given feed. Thus F_2 has 2 units of iron in one pound. Now suppose the cost of each feed in cents per pound is given in the following table.

F_1	F_2	F_3	F_4
2	3	2	3

A typical pig needs 5 units of iron, 8 of protein, 6 of folic acid, 7 of copper and 4 of calcium. (The units may change from nutrient to nutrient.) How many pounds of each feed per pig should the pig farmer use in order to minimize his cost?

His problem is to minimize $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ subject to the constraints

$$\begin{aligned}
x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\
5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\
x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\
2x_1 + x_2 + x_3 + x_4 &\geq 7, \\
x_1 + x_2 + x_3 + x_4 &\geq 4.
\end{aligned}$$

where each $x_i \geq 0$. Add in the slack variables,

$$\begin{aligned}
x_1 + 2x_2 + x_3 + 3x_4 - x_5 &= 5 \\
5x_1 + 3x_2 + 2x_3 + x_4 - x_6 &= 8 \\
x_1 + 2x_2 + 2x_3 + x_4 - x_7 &= 6 \\
2x_1 + x_2 + x_3 + x_4 - x_8 &= 7 \\
x_1 + x_2 + x_3 + x_4 - x_9 &= 4
\end{aligned}$$

The augmented matrix for this system is

$$\begin{pmatrix}
1 & 2 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 5 \\
5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 8 \\
1 & 2 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 6 \\
2 & 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 7 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 4
\end{pmatrix}$$

How in the world can you find a basic feasible solution? Remember the simplex algorithm is designed to keep the entries in the right column nonnegative so you use this algorithm a few times till the obvious solution is a basic feasible solution.

Consider the first column. The pivot is the 5. Using the row operations described in the algorithm, you get

$$\begin{pmatrix} 0 & \frac{7}{5} & \frac{3}{5} & \frac{14}{5} & -1 & \frac{1}{5} & 0 & 0 & 0 & \frac{17}{5} \\ 1 & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & 0 & -\frac{1}{5} & 0 & 0 & 0 & \frac{8}{5} \\ 0 & \frac{7}{5} & \frac{8}{5} & \frac{4}{5} & 0 & \frac{1}{5} & -1 & 0 & 0 & \frac{22}{5} \\ 0 & -\frac{1}{5} & \frac{1}{5} & \frac{3}{5} & 0 & \frac{2}{5} & 0 & -1 & 0 & \frac{19}{5} \\ 0 & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 0 & \frac{1}{5} & 0 & 0 & -1 & \frac{12}{5} \end{pmatrix}$$

Now go to the second column. The pivot in this column is the $7/5$. This is in a different row than the pivot in the first column so I will use it to zero out everything below it. This will get rid of the zeros in the fifth column and introduce zeros in the second. This yields

$$\begin{pmatrix} 0 & 1 & \frac{3}{7} & 2 & -\frac{5}{7} & \frac{1}{7} & 0 & 0 & 0 & \frac{17}{7} \\ 1 & 0 & \frac{1}{7} & -1 & \frac{3}{7} & -\frac{2}{7} & 0 & 0 & 0 & \frac{1}{7} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & \frac{2}{7} & 1 & -\frac{1}{7} & \frac{3}{7} & 0 & -1 & 0 & \frac{30}{7} \\ 0 & 0 & \frac{3}{7} & 0 & \frac{2}{7} & \frac{1}{7} & 0 & 0 & -1 & \frac{10}{7} \end{pmatrix}$$

Now consider another column, this time the fourth. I will pick this one because it has some negative numbers in it so there are fewer entries to check in looking for a pivot. Unfortunately, the pivot is the top 2 and I don't want to pivot on this because it would destroy the zeros in the second column. Consider the fifth column. It is also not a good choice because the pivot is the second element from the top and this would destroy the zeros in the first column. Consider the sixth column. I can use either of the two bottom entries as the pivot. The matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & -2 & 3 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 3 & 0 & 2 & 1 & 0 & 0 & -7 & 10 \end{pmatrix}$$

Next consider the third column. The pivot is the 1 in the third row. This yields

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 0 & 6 & -1 & 1 & 3 & 0 & -7 & 7 \end{pmatrix}.$$

There are still 5 columns which consist entirely of zeros except for one entry. Four of them have that entry equal to 1 but one still has a -1 in it, the -1 being in the fourth column. I need to do the row operations on a non-simple column which has the pivot in the fourth row. Such a column is the second to the last. The pivot is the 3. The new matrix is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & \frac{28}{3} \end{pmatrix}. \tag{11.15}$$

Now the obvious basic solution is feasible. You let $x_4 = 0 = x_5 = x_7 = x_8$ and $x_1 = 8/3, x_2 = 2/3, x_3 = 1,$ and $x_6 = 28/3$. You don't need to worry too much about this. It is the above matrix which is desired. Now you can assemble the simplex tableau and begin the algorithm. Remember $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$. First add the row and column which deal with C. This yields

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ -2 & -3 & -2 & -3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{11.16}$$

Now you do row operations to keep the simple columns of 11.15 simple in 11.16. Of course you could permute the columns if you wanted but this is not necessary.

This yields the following for a simplex tableau. Now it is a matter of getting rid of the positive entries in the bottom row because you are trying to minimize.

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

The most positive of them is the $2/3$ and so I will apply the algorithm to this one first. The pivot is the $7/3$. After doing the row operation the next tableau is

$$\begin{pmatrix} 0 & \frac{3}{7} & 0 & 1 & -\frac{3}{7} & 0 & \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{2}{7} \\ 1 & -\frac{1}{7} & 0 & 0 & \frac{1}{7} & 0 & \frac{2}{7} & -\frac{5}{7} & 0 & 0 & \frac{18}{7} \\ 0 & \frac{6}{7} & 1 & 0 & \frac{1}{7} & 0 & -\frac{5}{7} & \frac{2}{7} & 0 & 0 & \frac{11}{7} \\ 0 & \frac{1}{7} & 0 & 0 & -\frac{1}{7} & 0 & -\frac{2}{7} & -\frac{2}{7} & 1 & 0 & \frac{3}{7} \\ 0 & -\frac{11}{7} & 0 & 0 & \frac{4}{7} & 1 & \frac{1}{7} & -\frac{20}{7} & 0 & 0 & \frac{58}{7} \\ 0 & -\frac{2}{7} & 0 & 0 & -\frac{5}{7} & 0 & -\frac{3}{7} & -\frac{3}{7} & 0 & 1 & \frac{64}{7} \end{pmatrix}$$

and you see that all the entries are negative and so the minimum is $64/7$ and it occurs when $x_1 = 18/7, x_2 = 0, x_3 = 11/7, x_4 = 2/7$.

There is no maximum for the above problem. However, I will pretend I don't know this and attempt to use the simplex algorithm. You set up the simiplex tableau the same way. Recall it is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

Now to maximize, you try to get rid of the negative entries in the bottom left row. The most negative entry is the -1 in the fifth column. The pivot is the 1 in the third row of this column. The new tableau is

$$\begin{pmatrix} 0 & 1 & 1 & \frac{1}{3} & 0 & 0 & -\frac{2}{3} & \frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{5}{3} & 0 & 1 & -\frac{1}{3} & -\frac{7}{3} & 0 & 0 & \frac{31}{3} \\ 0 & 0 & 1 & -\frac{4}{3} & 0 & 0 & -\frac{4}{3} & -\frac{1}{3} & 0 & 1 & \frac{31}{3} \end{pmatrix}.$$

Consider the fourth column. The pivot is the top $1/3$. The new tableau is

$$\begin{pmatrix} 0 & 3 & 3 & 1 & 0 & 0 & -2 & 1 & 0 & 0 & 5 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 6 & 7 & 0 & 1 & 0 & -5 & 2 & 0 & 0 & 11 \\ 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 2 \\ 0 & -5 & -4 & 0 & 0 & 1 & 3 & -4 & 0 & 0 & 2 \\ 0 & 4 & 5 & 0 & 0 & 0 & -4 & 1 & 0 & 1 & 17 \end{pmatrix}$$

There is still a negative in the bottom, the -4. The pivot in that column is the 3. The algorithm yields

$$\begin{pmatrix} 0 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 & \frac{2}{3} & 0 & -\frac{5}{3} & 0 & 0 & \frac{19}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & -\frac{7}{3} & \frac{1}{3} & 0 & 1 & \frac{5}{3} & 0 & -\frac{14}{3} & 0 & 0 & \frac{43}{3} \\ 0 & -\frac{2}{3} & -\frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & -\frac{4}{3} & 1 & 0 & \frac{8}{3} \\ 0 & -\frac{5}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & 1 & -\frac{4}{3} & 0 & 0 & \frac{2}{3} \\ 0 & -\frac{8}{3} & -\frac{1}{3} & 0 & 0 & \frac{4}{3} & 0 & -\frac{13}{3} & 0 & 1 & \frac{59}{3} \end{pmatrix}$$

Note how z keeps getting larger. Consider the column having the $-13/3$ in it. The pivot is the single positive entry, $1/3$. The next tableau is

$$\begin{pmatrix} 5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 8 \\ 3 & 2 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 1 \\ 14 & 7 & 5 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 19 \\ 4 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 4 \\ 4 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 2 \\ 13 & 6 & 4 & 0 & 0 & -3 & 0 & 0 & 0 & 1 & 24 \end{pmatrix}.$$

There is a column consisting of all negative entries. There is therefore, no maximum. Note also how there is no way to pick the pivot in that column.

Example 11.3.4 *Minimize* $z = x_1 - 3x_2 + x_3$ *subject to the constraints* $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ *and* $x_1 + 2x_2 + x_3 \leq 7$ *with all variables nonnegative.*

There exists an answer because the region defined by the constraints is closed and bounded. Adding in slack variables you get the following augmented matrix corresponding to the constraints.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ 1 & 1 & 3 & 0 & 0 & 1 & 0 & 8 \\ 1 & 2 & 1 & 0 & 0 & 0 & 1 & 7 \end{pmatrix}$$

Of course there is a problem with the obvious solution obtained by setting to zero all variables corresponding to a nonsimple column because of the simple column which has the -1 in it. Therefore, I will use the simplex algorithm to make this column non simple. The third column has the 1 in the second row as the pivot so I will use this column. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 5 \end{pmatrix} \tag{11.17}$$

and the obvious solution is feasible. Now it is time to assemble the simplex tableau. First add in the bottom row and second to last column corresponding to the equation for z. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ -1 & 3 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Next you need to zero out the entries in the bottom row which are below one of the simple columns in 11.17. This yields the simplex tableau

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

The desire is to minimize this so you need to get rid of the positive entries in the left bottom row. There is only one such entry, the 4. In that column the pivot is the 1 in the second row of this column. Thus the next tableau is

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 6 \\ -1 & 0 & -1 & 0 & 2 & 0 & 1 & 0 & 3 \\ -4 & 0 & -4 & 0 & 3 & 0 & 0 & 1 & -6 \end{pmatrix}$$

There is still a positive number there, the 3. The pivot in this column is the 2. Apply the algorithm again. This yields

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & -\frac{1}{2} & 0 & \frac{13}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{7}{2} \\ \frac{1}{2} & 0 & \frac{5}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & \frac{9}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{5}{2} & 0 & -\frac{5}{2} & 0 & 0 & 0 & -\frac{3}{2} & 1 & -\frac{21}{2} \end{pmatrix}.$$

Now all the entries in the left bottom row are nonpositive so the process has stopped. The minimum is $-21/2$. It occurs when $x_1 = 0$, $x_2 = 7/2, x_3 = 0$.

Now consider the same problem but change the word, minimize to the word, maximize.

Example 11.3.5 *Maximize* $z = x_1 - 3x_2 + x_3$ *subject to the constraints* $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ *and* $x_1 + 2x_2 + x_3 \leq 7$ *with all variables nonnegative.*

The first part of it is the same. You wind up with the same simplex tableau,

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}$$

but this time, you apply the algorithm to get rid of the negative entries in the left bottom row. There is a -1. Use this column. The pivot is the 3. The next tableau is

$$\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & 0 & 1 & 0 & -\frac{1}{3} & 0 & 0 & \frac{22}{3} \\ \frac{1}{3} & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{8}{3} \\ -\frac{2}{3} & -\frac{2}{3} & 0 & 0 & 1 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{5}{3} & 0 & 0 & 0 & -\frac{1}{3} & 1 & 0 & \frac{13}{3} \\ -\frac{2}{3} & \frac{10}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 1 & \frac{8}{3} \end{pmatrix}$$

There is still a negative entry, the $-2/3$. This will be the new pivot column. The pivot is the $2/3$ on the fourth row. This yields

$$\begin{pmatrix} 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 3 \\ 0 & -\frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 1 & \frac{5}{2} & 0 & 0 & 0 & -\frac{1}{2} & \frac{3}{2} & 0 & \frac{13}{2} \\ 0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 7 \end{pmatrix}$$

and the process stops. The maximum for z is 7 and it occurs when $x_1 = 13/2, x_2 = 0, x_3 = 1/2$.

11.4 FINDING A BASIC FEASIBLE SOLUTION

By now it should be fairly clear that finding a basic feasible solution can create considerable difficulty. Indeed, given a system of linear inequalities along with the requirement that each variable be nonnegative, do there even exist points satisfying all these inequalities? If you have many variables, you can't answer this by drawing a picture. Is there some other way to do this which is more systematic than what was presented above? The answer is yes. It is called the method of artificial variables. I will illustrate this method with an example.

Example 11.4.1 *Find a basic feasible solution to the system* $2x_1 + x_2 - x_3 \geq 3, x_1 + x_2 + x_3 \geq 2, x_1 + x_2 + x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

If you write the appropriate augmented matrix with the slack variables,

$$\left(\begin{array}{cccccc|c} 2 & 1 & -1 & -1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 7 \end{array}\right) \tag{11.18}$$

The obvious solution is not feasible. This is why it would be hard to get started with the simplex method. What is the problem? It is those -1 entries in the fourth and fifth columns. To get around this, you add in artificial variables to get an augmented matrix of the form

$$\left(\begin{array}{cccccccc|c} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 7 \end{array}\right) \tag{11.19}$$

Thus the variables are $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$. Suppose you can find a feasible solution to the system of equations represented by the above augmented matrix. Thus all variables are non-negative. Suppose also that it can be done in such a way that x_8 and x_7 happen to be 0. Then it will follow that $x_1, \cdots, x_6$ is a feasible solution for 11.18. Conversely, if you can find a feasible solution for 11.18, then letting x_7 and x_8 both equal zero, you have obtained a feasible solution to 11.19. Since all variables are nonnegative, x_7 and x_8 both equalling zero is equivalent to saying the minimum of $z = x_7 + x_8$ subject to the constraints represented by the above augmented matrix equals zero. This has proved the following simple observation.

Observation 11.4.2 *There exists a feasible solution to the constraints represented by the augmented matrix of 11.18 and* $\mathbf{x} \geq \mathbf{0}$ *if and only if the minimum of* $x_7 + x_8$ *subject to the constraints of 11.19 and* $\mathbf{x} \geq \mathbf{0}$ *exists and equals* $\mathbf{0}$.

Of course a similar observation would hold in other similar situations. Now the point of all this is that it is trivial to see a feasible solution to 11.19, namely $x_6 = 7, x_7 = 3, x_8 = 2$ and all the other variables may be set to equal zero. Therefore, it is easy to find an initial simplex tableau for the minimization problem just described. First add the column and row for z

$$\left(\begin{array}{ccccccccc|c} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{array}\right)$$

Next it is necessary to make the last two columns on the bottom left row into simple columns. Performing the row operation, this yields an initial simplex tableau,

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 3 & 2 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 5 \end{pmatrix}$$

Now the algorithm involves getting rid of the positive entries on the left bottom row. Begin with the first column. The pivot is the 2. An application of the simplex algorithm yields the new tableau

$$\begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{1}{2} & 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0 & 1 & -\frac{1}{2} & 0 & 0 & \frac{11}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{3}{2} & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

Now go to the third column. The pivot is the $3/2$ in the second row. An application of the simplex algorithm yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix} \tag{11.20}$$

and you see there are only non-positive numbers on the bottom left column so the process stops and yields $\mathbf{0}$ for the minimum of $z = x_7 + x_8$. As for the other variables, $x_1 = 5/3, x_2 = 0, x_3 = 1/3, x_4 = 0, x_5 = 0, x_6 = 5$. Now as explained in the above observation, this is a basic feasible solution for the original system 11.18.

Now consider a maximization problem associated with the above constraints.

Example 11.4.3 *Maximize* $x_1 - x_2 + 2x_3$ *subject to the constraints,* $2x_1 + x_2 - x_3 \geq 3, x_1 + x_2 + x_3 \geq 2, x_1 + x_2 + x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

From 11.20 you can immediately assemble an initial simplex tableau. You begin with the first 6 columns and top 3 rows in 11.20. Then add in the column and row for z. This yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ -1 & 1 & -2 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and you first do row operations to make the first and third columns simple columns. Thus the next simplex tableau is

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & -\frac{5}{3} & 0 & 1 & \frac{7}{3} \end{pmatrix}$$

You are trying to get rid of negative entries in the bottom left row. There is only one, the $-5/3$. The pivot is the 1. The next simplex tableau is then

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{10}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{11}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & 0 & \frac{5}{3} & 1 & \frac{32}{3} \end{pmatrix}$$

and so the maximum value of z is $32/3$ and it occurs when $x_1 = 10/3, x_2 = 0$ and $x_3 = 11/3$.

11.5 DUALITY

You can solve minimization problems by solving maximization problems. You can also go the other direction and solve maximization problems by minimization problems. Sometimes this makes things much easier. To be more specific, the two problems to be considered are

A) Minimize $z = \mathbf{cx}$ subject to $\mathbf{x} \geq \mathbf{0}$ and $A\mathbf{x} \geq \mathbf{b}$ and
B) Maximize $w = \mathbf{yb}$ such that $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}A \leq \mathbf{c}$,

$$\left(\text{equivalently } A^T\mathbf{y}^T \geq \mathbf{c}^T \text{ and } w = \mathbf{b}^T\mathbf{y}^T\right).$$

In these problems it is assumed A is an $m \times p$ matrix.

I will show how a solution of the first yields a solution of the second and then show how a solution of the second yields a solution of the first. The problems, $A.$) and $B.$) are called dual problems.

Lemma 11.5.1 *Let* $\mathbf{x}$ *be a solution of the inequalities of* $A.$) *and let* $\mathbf{y}$ *be a solution of the inequalities of* $B.$) *Then*

$$\mathbf{cx} \geq \mathbf{yb}.$$

and if equality holds in the above, then $\mathbf{x}$ *is the solution to* $A.$) *and* $\mathbf{y}$ *is a solution to* $B.$)

Proof: This follows immediately. Since $\mathbf{c} \geq \mathbf{y}A$,

$$\mathbf{cx} \geq \mathbf{y}A\mathbf{x} \geq \mathbf{yb}.$$

It follows from this lemma that if $\mathbf{y}$ satisfies the inequalities of $B.$) and $\mathbf{x}$ satisfies the inequalities of $A.$) then if equality holds in the above lemma, it must be that $\mathbf{x}$ is a solution of $A.$) and $\mathbf{y}$ is a solution of $B.$). ■

Now recall that to solve either of these problems using the simplex method, you first add in slack variables. Denote by $\mathbf{x}'$ and $\mathbf{y}'$ the enlarged list of variables. Thus $\mathbf{x}'$ has at least entries and so does $\mathbf{y}'$ and the inequalities involving A were replaced by equalities whose augmented matrices were of the form

$$\begin{pmatrix} A & -I & \mathbf{b} \end{pmatrix}, \text{ and } \begin{pmatrix} A^T & I & \mathbf{c}^T \end{pmatrix}$$

Then you included the row and column for z and w to obtain

$$\begin{pmatrix} A & -I & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & \mathbf{0} & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} A^T & I & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}^T & \mathbf{0} & 1 & 0 \end{pmatrix}. \tag{11.21}$$

Then the problems have basic feasible solutions if it is possible to permute the first $p+m$ columns in the above two matrices and obtain matrices of the form

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} B_1 & F_1 & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}_{B_1}^T & -\mathbf{b}_{F_1}^T & 1 & 0 \end{pmatrix} \tag{11.22}$$

where B, B_1 are invertible $m \times m$ and $p \times p$ matrices and denoting the variables associated with these columns by $\mathbf{x}_B, \mathbf{y}_B$ and those variables associated with F or F_1 by $\mathbf{x}_F$ and $\mathbf{y}_F$, it follows that letting $B\mathbf{x}_B = \mathbf{b}$ and $\mathbf{x}_F = \mathbf{0}$, the resulting vector $\mathbf{x}'$ is a solution to $\mathbf{x}' \geq \mathbf{0}$ and $\begin{pmatrix} A & -I \end{pmatrix} \mathbf{x}' = \mathbf{b}$ with similar constraints holding for $\mathbf{y}'$. In other words, it is possible to obtain simplex tableaus,

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{pmatrix}, \begin{pmatrix} I & B_1^{-1}F_1 & \mathbf{0} & B_1^{-1}\mathbf{c}^T \\ \mathbf{0} & \mathbf{b}_{B_1}^T B_1^{-1}F - \mathbf{b}_{F_1}^T & 1 & \mathbf{b}_{B_1}^T B_1^{-1}\mathbf{c}^T \end{pmatrix} \tag{11.23}$$

Similar considerations apply to the second problem. Thus as just described, a basic feasible solution is one which determines a simplex tableau like the above in which you get a feasible solution by setting all but the first m variables equal to zero. The simplex algorithm takes you from one basic feasible solution to another till eventually, if there is no degeneracy, you obtain a basic feasible solution which yields the solution of the problem of interest.

Theorem 11.5.2 *Suppose there exists a solution,* $\mathbf{x}$ *to* $A.)$ *where* $\mathbf{x}$ *is a basic feasible solution of the inequalities of* $A.)$. *Then there exists a solution,* $\mathbf{y}$ *to* $B.)$ *and* $\mathbf{cx} = \mathbf{by}$. *It is also possible to find* $\mathbf{y}$ *from* $\mathbf{x}$ *using a simple formula.*

Proof: Since the solution to $A.)$ is basic and feasible, there exists a simplex tableau like 11.23 such that $\mathbf{x}'$ can be split into $\mathbf{x}_B$ and $\mathbf{x}_F$ such that $\mathbf{x}_F = \mathbf{0}$ and $\mathbf{x}_B = B^{-1}\mathbf{b}$. Now since it is a minimizer, it follows $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ and the minimum value for $\mathbf{cx}$ is $\mathbf{c}_B B^{-1}\mathbf{b}$. Stating this again, $\mathbf{cx} = \mathbf{c}_B B^{-1}\mathbf{b}$. Is it possible you can take $\mathbf{y} = \mathbf{c}_B B^{-1}$? From Lemma 11.5.1 this will be so if $\mathbf{c}_B B^{-1}$ solves the constraints of problem $B.)$. Is $\mathbf{c}_B B^{-1} \geq 0$? Is $\mathbf{c}_B B^{-1}A \leq \mathbf{c}$? These two conditions are satisfied if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} A & -I \end{pmatrix} \leq \begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$. Referring to the process of permuting the columns of the first augmented matrix of 11.21 to get 11.22 and doing the same permutations on the columns of $\begin{pmatrix} A & -I \end{pmatrix}$ and $\begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$, and the desired inequality holds if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} B & F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ which is equivalent to saying $\begin{pmatrix} \mathbf{c}_B & \mathbf{c}_B B^{-1}F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ and this is true because $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ due to the assumption that $\mathbf{x}$ is a minimizer. The simple formula is just

$$\mathbf{y} = \mathbf{c}_B B^{-1}. \blacksquare$$

The proof of the following corollary is similar.

Corollary 11.5.3. *Suppose there exists a solution,* $\mathbf{y}$ *to* $B.)$ *where* $\mathbf{y}$ *is a basic feasible solution of the inequalities of* $B.)$*. Then there exists a solution,* $\mathbf{x}$ *to* $A.)$ *and* $\mathbf{cx} = \mathbf{by}$*. It is also possible to find* $\mathbf{x}$ *from* $\mathbf{y}$ *using a simple formula. In this case, and referring to 11.23, the simple formula is* $\mathbf{x} = B_1^{-T}\mathbf{b}_{B_1}$.

As an example, consider the pig farmers problem. The main difficulty in this problem was finding an initial simplex tableau. Now consider the following example and marvel at how all the difficulties disappear.

Example 11.5.4 *minimize* $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ *subject to the constraints*

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\ 5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\ x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\ 2x_1 + x_2 + x_3 + x_4 &\geq 7, \\ x_1 + x_2 + x_3 + x_4 &\geq 4. \end{aligned}$$

where each $x_i \geq 0$.

Here the dual problem is to maximize $w = 5y_1 + 8y_2 + 6y_3 + 7y_4 + 4y_5$ subject to the constraints

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 \\ 2 & 3 & 2 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \end{pmatrix}.$$

Adding in slack variables, these inequalities are equivalent to the system of equations whose augmented matrix is

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

Now the obvious solution is feasible so there is no hunting for an initial obvious feasible solution required. Now add in the row and column for w. This yields

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

It is a maximization problem so you want to eliminate the negatives in the bottom left row. Pick the column having the one which is most negative, the -8. The pivot is the top 5. Then apply the simplex algorithm to obtain

$$\begin{pmatrix} \frac{1}{5} & 1 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & \frac{2}{5} \\ \frac{7}{5} & 0 & \frac{7}{5} & -\frac{1}{5} & \frac{2}{5} & -\frac{3}{5} & 1 & 0 & 0 & 0 & \frac{9}{5} \\ \frac{3}{5} & 0 & \frac{8}{5} & \frac{1}{5} & \frac{3}{5} & -\frac{2}{5} & 0 & 1 & 0 & 0 & \frac{6}{5} \\ \frac{14}{5} & 0 & \frac{4}{5} & \frac{3}{5} & \frac{4}{5} & -\frac{1}{5} & 0 & 0 & 1 & 0 & \frac{13}{5} \\ -\frac{17}{5} & 0 & -\frac{22}{5} & -\frac{19}{5} & -\frac{12}{5} & \frac{8}{5} & 0 & 0 & 0 & 1 & \frac{16}{5} \end{pmatrix}.$$

There are still negative entries in the bottom left row. Do the simplex algorithm to the column which has the $-\frac{22}{5}$. The pivot is the $\frac{8}{5}$. This yields

$$\begin{pmatrix} \frac{1}{8} & 1 & 0 & \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & 0 & -\frac{1}{8} & 0 & 0 & \frac{1}{4} \\ \frac{7}{8} & 0 & 0 & -\frac{3}{8} & -\frac{1}{8} & -\frac{1}{4} & 1 & -\frac{7}{8} & 0 & 0 & \frac{3}{4} \\ \frac{3}{8} & 0 & 1 & \frac{1}{8} & \frac{3}{8} & -\frac{1}{4} & 0 & \frac{5}{8} & 0 & 0 & \frac{3}{4} \\ \frac{5}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 & 2 \\ -\frac{7}{4} & 0 & 0 & -\frac{13}{4} & -\frac{3}{4} & \frac{1}{2} & 0 & \frac{11}{4} & 0 & 1 & \frac{13}{2} \end{pmatrix}$$

and there are still negative numbers. Pick the column which has the $-13/4$. The pivot is the $3/8$ in the top. This yields

$$\begin{pmatrix} \frac{1}{3} & \frac{8}{3} & 0 & 1 & \frac{1}{3} & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ \frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \frac{2}{3} \\ \frac{7}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & 1 & 0 & \frac{5}{3} \\ -\frac{2}{3} & \frac{26}{3} & 0 & 0 & \frac{1}{3} & \frac{8}{3} & 0 & \frac{5}{3} & 0 & 1 & \frac{26}{3} \end{pmatrix}$$

which has only one negative entry on the bottom left. The pivot for this first column is the $\frac{7}{3}$. The next tableau is

$$\begin{pmatrix} 0 & \frac{20}{7} & 0 & 1 & \frac{2}{7} & \frac{5}{7} & 0 & -\frac{2}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 0 & \frac{11}{7} & 0 & 0 & -\frac{1}{7} & \frac{1}{7} & 1 & -\frac{6}{7} & -\frac{3}{7} & 0 & \frac{2}{7} \\ 0 & -\frac{1}{7} & 1 & 0 & \frac{2}{7} & -\frac{2}{7} & 0 & \frac{5}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 1 & -\frac{4}{7} & 0 & 0 & \frac{1}{7} & -\frac{1}{7} & 0 & -\frac{1}{7} & \frac{3}{7} & 0 & \frac{5}{7} \\ 0 & \frac{58}{7} & 0 & 0 & \frac{3}{7} & \frac{18}{7} & 0 & \frac{11}{7} & \frac{2}{7} & 1 & \frac{64}{7} \end{pmatrix}$$

and all the entries in the left bottom row are nonnegative so the answer is $64/7$. This is the same as obtained before. So what values for are needed? Here the basic variables are y_1, y_3, y_4, y_7. Consider the original augmented matrix, one step before the simplex tableau.

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Permute the columns to put the columns associated with these basic variables first. Thus

$$\begin{pmatrix} 1 & 1 & 2 & 0 & 5 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 & 3 \\ 1 & 2 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 3 \\ -5 & -6 & -7 & 0 & -8 & -4 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The matrix B is

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

and so B^{-T} equals

$$\begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix}$$

Also $\mathbf{b}_B^T = \begin{pmatrix} 5 & 6 & 7 & 0 \end{pmatrix}$ and so from Corollary 11.5.3,

$$\mathbf{x} = \begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix} \begin{pmatrix} 5 \\ 6 \\ 7 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{18}{7} \\ 0 \\ \frac{11}{7} \\ \frac{2}{7} \end{pmatrix}$$

which agrees with the original way of doing the problem.

Two good books which give more discussion of linear programming are Strang [17] and Nobel and Daniels [14]. Also listed in these books are other references which may prove useful if you are interested in seeing more on these topics. There is a great deal more which can be said about linear programming.

11.6 EXERCISES

1. Maximize and minimize $z = x_1 - 2x_2 + x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 10,\ x_1 + \ x_2 + x_3 \geq 2$ and $x_1 + 2x_2 + x_3 \leq 7$ if possible. All variables are nonnegative.
2. Maximize and minimize the following if possible. All variables are nonnegative.

   a) $z = x_1 - 2x_2$ subject to the constraints $x_1 + x_2 + x_3 \leq 10,\ x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$.
   b) $z = x_1 - 2x_2 - 3x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 8,\ x_1 + x_2 + 3x_3 \geq 1$, and $x_1 + x_2 + x_3 \leq 7$.

c) $z = 2x_1 + x_2$ subject to the constraints $x_1 - x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$.

d) $z = x_1 + 2x_2$ subject to the constraints $x_1 - x_2 + x_3 \leq 10$, $x_1 + x_2 + x_3 \geq 1$, and $x_1 + 2x_2 + x_3 \leq 7$.

3. Consider contradictory constraints, $x_1 + x_2 \geq 12$ and $x_1 + 2x_2 \leq 5, x_1 \geq 0, x_2 \geq 0$. You know these two contradict but show they contradict using the simplex algorithm.
4. Find a solution to the following inequalities for $x, y \geq 0$ if it is possible to do so. If it is not possible, prove it is not possible.

a) $\begin{array}{c} 6x + 3y \geq 4 \\ 8x + 4y \leq 5 \end{array}$

b) $\begin{array}{c} 6x_1 + 4x_3 \leq 11 \\ 5x_1 + 4x_2 + 4x_3 \geq 8 \\ 6x_1 + 6x_2 + 5x_3 \leq 11 \end{array}$

c) $\begin{array}{c} 6x_1 + 4x_3 \leq 11 \\ 5x_1 + 4x_2 + 4x_3 \geq 9 \\ 6x_1 + 6x_2 + 5x_3 \leq 9 \end{array}$

d) $\begin{array}{c} x_1 - x_2 + x_3 \leq 2 \\ x_1 + 2x_2 \geq 4 \\ 3x_1 + 2x_3 \leq 7 \end{array}$

e) $\begin{array}{c} 5x_1 - 2x_2 + 4x_3 \leq 1 \\ 6x_1 - 3x_2 + 5x_3 \geq 2 \\ 5x_1 - 2x_2 + 4x_3 \leq 5 \end{array}$

5. Minimize $z = x_1 + x_2$ subject to $x_1 + x_2 \geq 2$, $x_1 + 3x_2 \leq 20$, $x_1 + x_2 \leq 18$. Change to a maximization problem and solve as follows: $y_i = M - x_i$. Let Formulate in terms of y_1, y_2.

12 SPECTRAL THEORY

12.1 EIGENVALUES AND EIGENVECTORS OF A MATRIX

Spectral Theory refers to the study of eigenvalues and eigenvectors of a matrix. It is of fundamental importance in many areas. Row operations will no longer be the solution to all issues.

12.1.1 DEFINITION OF EIGENVECTORS AND EIGENVALUES

In this section, $\mathbb{F} = \mathbb{C}$. In fact, it is assumed that the field is one for which eigenvalues exist and this will mean $\mathbb{C}$.

To illustrate the idea behind what will be discussed, consider the following example.

Example 12.1.1 *Here is a matrix.*

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix}.$$

Multiply this matrix by the vector $\begin{pmatrix} 5 & -4 & 3 \end{pmatrix}^T$ *and see what happens. Then multiply it by* $\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$ *and see what happens. Does this matrix act this way for some other vector?*

First

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix} = \begin{pmatrix} -50 \\ -40 \\ 30 \end{pmatrix} = 10 \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix}.$$

Next

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

When you multiply the first vector by the given matrix, it stretched the vector, multiplying it by When you multiplied the matrix by the second vector it sent it to the zero vector. Now consider

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ 38 \\ -11 \end{pmatrix}.$$

In this case, multiplication by the matrix did not result in merely multiplying the vector by a number.

In the above example, the first two vectors were called eigenvectors and the numbers, 10 and 0 are called eigenvalues. Not every number is an eigenvalue and not every vector is an eigenvector. When you have a **nonzero** vector which, when multiplied by a matrix results in another vector which is parallel to the first or equal to **0,** this vector is called an eigenvector of the matrix. This is the meaning when the vectors are in $\mathbb{R}^n$. Things are less apparent geometrically when the vectors are in $\mathbb{C}^n$. The precise definition in all cases follows.

Definition 12.1.2 *Let M be an $n \times n$ matrix and let $\mathbf{x} \in \mathbb{C}^n$ be a* ***nonzero vector*** *for which*

$$M\mathbf{x} = \lambda \mathbf{x} \tag{12.1}$$

for some scalar λ. Then $\mathbf{x}$ is called an ***eigenvector*** *and λ is called an* ***eigenvalue (characteristic value)*** *of the matrix M.*

Note: Eigenvectors are never equal to zero!

The set of all eigenvalues of an $n \times n$ matrix M, is denoted by $\sigma(M)$ and is referred to as the ***spectrum*** *of M.*

The eigenvectors of a matrix M are those vectors, $\mathbf{x}$ for which multiplication by M results in a vector in the same direction or opposite direction to $\mathbf{x}$. Since the zero vector $\mathbf{0}$ has no direction this would make no sense for the zero vector. As noted above, $\mathbf{0}$ is never allowed to be an eigenvector. How can eigenvectors be identified? Suppose $\mathbf{x}$ satisfies 12.1. Then

$$(M - \lambda I)\mathbf{x} = \mathbf{0}$$

for some $\mathbf{x} \neq \mathbf{0}$. (Equivalently, you could write $(\lambda I - M)\mathbf{x} = \mathbf{0}$.) Sometimes we will use

$$(\lambda I - M)\mathbf{x} = \mathbf{0}$$

and sometimes $(M - \lambda I)\mathbf{x} = \mathbf{0}$. It makes absolutely no difference and you should use whichever you like better. Therefore, the matrix $M - \lambda I$ cannot have an inverse because if it did, the equation could be solved,

$$\mathbf{x} = \left((M - \lambda I)^{-1}(M - \lambda I)\right)\mathbf{x} = (M - \lambda I)^{-1}((M - \lambda I)\mathbf{x}) = (M - \lambda I)^{-1}\mathbf{0} = \mathbf{0},$$

and this would require $\mathbf{x} = \mathbf{0}$, contrary to the requirement that $\mathbf{x} \neq \mathbf{0}$. By Theorem 6.2.1 on Page 171,

$$\det(M - \lambda I) = 0. \tag{12.2}$$

(Equivalently you could write $\det(M - \lambda I) = 0$.) The expression, $\det(\lambda I - M)$ or equivalently, $\det(\lambda I - M)$ is a polynomial called the **characteristic polynomial** and the above equation is called the characteristic equation.For M an $n \times n$ matrix, it follows from the theorem on expanding a matrix by its cofactor that $\det(\lambda I - M)$ is a polynomial of degree n. As such, the equation 12.2 has a solution, $\lambda \in \mathbb{C}$ by the fundamental theorem of algebra. Is it actually an eigenvalue? The answer is yes, and this follows from Observation 9.2.7 on Page 283 along with Theorem 6.2.1 on Page 171. Since $\det(M - \lambda I) = 0$ the matrix $\det(\lambda I - M)$ cannot be one to one and so there exists a nonzero $\mathbf{x}$ vector such that $(M - \lambda I)\mathbf{x} = \mathbf{0}$. This proves the following corollary.

Corollary 12.1.3 *Let M be an $n \times n$ matrix and* $\det(M - \lambda I) = 0$*. Then there exists a nonzero vector* $\mathbf{x} \in \mathbb{C}^n$ *such that* $(M - \lambda I)\mathbf{x} = \mathbf{0}$.

12.1.2 FINDING EIGENVECTORS AND EIGENVALUES

As an example, consider the following.

Example 12.1.4 *Find the eigenvalues and eigenvectors for the matrix*

$$A = \begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix}.$$

You first need to identify the eigenvalues. Recall this requires the solution of the equation $\det(A - \lambda I) = 0$. In this case this equation is

$$\det\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0$$

When you expand this determinant and simplify, you find the equation you need to solve is

$$(\lambda - 5)\left(\lambda^2 - 20\lambda + 100\right) = 0$$

and so the eigenvalues are $5, 10, 10$. We have listed twice because it is a zero of multiplicity two due to

$$\lambda^2 - 20\lambda + 100 = (\lambda - 10)^2 .$$

Having found the eigenvalues, it only remains to find the eigenvectors. First find the eigenvectors for $\lambda = 5$. As explained above, this requires you to solve the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 5\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

That is you need to find the solution to

$$\begin{pmatrix} 0 & -10 & -5 \\ 2 & 9 & 2 \\ -4 & -8 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

By now this is an old problem. You set up the augmented matrix and row reduce to get the solution. Thus the matrix you must row reduce is

$$\begin{pmatrix} 0 & -10 & -5 & | & 0 \\ 2 & 9 & 2 & | & 0 \\ -4 & -8 & 1 & | & 0 \end{pmatrix}. \tag{12.3}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -\frac{5}{4} & | & 0 \\ 0 & 1 & \frac{1}{2} & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

and so the solution is any vector of the form

$$\begin{pmatrix} \frac{5}{4}t \\ \frac{-1}{2}t \\ t \end{pmatrix} = t\begin{pmatrix} \frac{5}{4} \\ \frac{-1}{2} \\ 1 \end{pmatrix}$$

where $t \in \mathbb{F}$. You would obtain the same collection of vectors if you replaced t with $4t$. Thus a simpler description for the solutions to this system of equations whose augmented matrix is in 12.3 is

$$t\begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} \tag{12.4}$$

where $t \in \mathbb{F}$. Now you need to remember that you can't take $t = 0$ because this would result in the zero vector and

Eigenvectors are never equal to zero!

Other than this value, every other choice of z in 12.4 results in an eigenvector. It is a good idea to check your work! To do so, we will take the original matrix and multiply by this vector and see if we get 5 times this vector.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix}\begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 25 \\ -10 \\ 20 \end{pmatrix} = 5\begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix}$$

so it appears this is correct. Always check your work on these problems if you care about getting the answer right.

The parameter, t is sometimes called a **free variable.** The set of vectors in 12.4 is called the **eigenspace** and it is defined by $(A - \lambda I)$. You should observe that in this case the eigenspace has dimension because the eigenspace is the span of a single vector. In general, you obtain the solution from the row echelon form and the number of different free variables gives you the dimension of the eigenspace. Just remember that not every vector in the eigenspace is an eigenvector. The vector 0 is not an eigenvector although it is in the eigenspace because

Eigenvectors are never equal to zero!

Next consider the eigenvectors for $\lambda = 10$. These vectors are solutions to the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 10\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

That is you must find the solutions to

$$\begin{pmatrix} -5 & -10 & -5 \\ 2 & 4 & 2 \\ -4 & -8 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which reduces to consideration of the augmented matrix

$$\left(\begin{array}{ccc|c} -5 & -10 & -5 & 0 \\ 2 & 4 & 2 & 0 \\ -4 & -8 & -4 & 0 \end{array}\right)$$

The row reduced echelon form for this matrix is

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the eigenvectors are of the form

$$\begin{pmatrix} -2s-t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

You can't pick t and s both equal to zero because this would result in the zero vector and

Eigenvectors are never equal to zero!

However, every other choice t of and s does result in an eigenvector for the eigenvalue $\lambda = 10$. As in the case for $\lambda = 5$ you should check your work if you care about getting it right.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix} = 10\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

so it worked. The other vector will also work. Check it.

12.2.3 A WARNING

The above example shows how to find eigenvectors and eigenvalues algebraically. You may have noticed it is a bit long. Sometimes students try to first row reduce the matrix before looking for eigenvalues. This is a $\boxed{\text{terrible idea}}$ because row operations destroy the eigenvalues. The eigenvalue problem is really not about row operations.

The general eigenvalue problem is the hardest problem in algebra and people still do research on ways to find eigenvalues and their eigenvectors. If you are doing anything which would yield a way to find eigenvalues and eigenvectors for general matrices without too much trouble, the thing you are doing will certainly be wrong. The problems you will see in this book are not too hard because they are cooked up to be easy. General methods to compute eigenvalues and eigenvectors numerically are presented later. These methods work even when the problem is not cooked up to be easy.

Notwithstanding the above discouraging observations, one can sometimes simplify the matrix first before searching for the eigenvalues in other ways.

Lemma 12.1.5 *Let $A = S^{-1}BS$ where A, B are $n \times n$ matrices. Then A, B have the same eigenvalues.*

Proof: Say $A\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$. Then

$$S^{-1}BS\mathbf{x} = \lambda\mathbf{x} \text{ and so } BS\mathbf{x} = \lambda S\mathbf{x}.$$

Since S is one to one, $S\mathbf{x} \neq \mathbf{0}$. Thus if is an eigenvalue for A then it is also an eigenvalue for B. The other direction is similar. ■

Note that from the proof of the lemma, the eigenvectors for A, B are different.

One can now attempt to simplify a matrix before looking for the eigenvalues by using elementary matrices for S. This is illustrated in the following example.

Example 12.1.6 *Find the eigenvalues for the matrix*

$$\begin{pmatrix} 33 & 105 & 105 \\ 10 & 28 & 30 \\ -20 & -60 & -62 \end{pmatrix}$$

It has big numbers. Use the row operation of adding two times the second row to the bottom and multiply by the inverse of the elementary matrix which does this row operation as illustrated.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 33 & 105 & 105 \\ 10 & 28 & 30 \\ -20 & -60 & -62 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{pmatrix}$$

This one has the same eigenvalues as the first matrix but is of course much easier to work with. Next, do the same sort of thing

$$\begin{pmatrix} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 15 \\ 10 & -2 & 30 \\ 0 & 0 & -2 \end{pmatrix} \tag{12.5}$$

At this point, you can get the eigenvalues easily. This last matrix has the same eigenvalues as the first. Thus the eigenvalues are obtained by solving

$$(-2 - \lambda)(-2 - \lambda)(3 - \lambda) = 0,$$

and so the eigenvalues of the original matrix are $-2, -2, 3$.

At this point, you go back to the **original matrix** A, form $A - \lambda I$, and then the problem from here on does reduce to row operations. In general, if you are so fortunate as to find the eigenvalues as in the above example, then finding the eigenvectors does reduce to row operations and this part of the problem is easy.

However, finding the eigenvalues along with the eigenvectors is anything but easy because for an $n \times n$ matrix A, it involves solving a polynomial equation of degree n. If you only find a good approximation to the eigenvalue, it won't work. It either is or is not an eigenvalue and if it is not, the only solution to the equation, $(A - \lambda I)\,\mathbf{x} = \mathbf{0}$ will be the zero solution as explained above and

Eigenvectors are never equal to zero!

Another thing worth noting is that when you multiply on the right by an elementary operation, you are merely doing the column operation defined by the elementary matrix. In 12.5 multiplication by the elementary matrix on the right merely involves taking three times the first column and adding to the second. Thus, without referring to the elementary matrices, the transition to the new matrix in 12.5 can be illustrated by

$$\begin{pmatrix} 33 & -105 & 105 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{pmatrix} \to \begin{pmatrix} 3 & -9 & 15 \\ 10 & -32 & 30 \\ 0 & 0 & -2 \end{pmatrix} \to \begin{pmatrix} 3 & 0 & 15 \\ 10 & -2 & 30 \\ 0 & 0 & -2 \end{pmatrix}$$

Here is another example.

Example 12.1.7 *Let*

$$A = \begin{pmatrix} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix}$$

First find the eigenvalues. If you like, you could use the above technique to simplify the matrix, obtaining one which has the same eigenvalues, but since the numbers are not large, it is probably better to just expand the determinant without any tricks.

$$\det\left(\begin{pmatrix} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0$$

This reduces to $\lambda^3 - 6\lambda^2 + 8\lambda = 0$ and the solutions are 0, 2, and 4.

0 Can be an Eigenvalue!

Now find the eigenvectors. For $\lambda = 0$ the augmented matrix for finding the solutions is

$$\left(\begin{array}{ccc|c} 2 & 2 & -2 & 0 \\ 1 & 3 & -1 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right)$$

and the row reduced echelon form is

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

Therefore, the eigenvectors are of the form $t\begin{pmatrix} 1 & 0 & 1 \end{pmatrix}^T$ where $t \neq 0$.

Next find the eigenvectors for $\lambda = 2$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c} -2 & 2 & -2 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & 1 & -3 & 0 \end{array}\right)$$

and the row reduced echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so the eigenvectors are of the $t\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}^T$ form where $t \neq 0$.

Finally find the eigenvectors for $\lambda = 4$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c} -2 & 2 & -2 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & 1 & -3 & 0 \end{array}\right)$$

and the row reduced echelon form is

$$\left(\begin{array}{cccc} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors are of the form $t\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}^T$ where $t \neq 0$.

12.1.4 TRIANGULAR MATRICES

Although it is usually hard to solve the eigenvalue problem, there is a kind of matrix for which this is not the case. These are the upper or lower triangular matrices. I will illustrate by examples.

Example 12.1.8 *Let* $A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 4 & 7 \\ 0 & 0 & 6 \end{pmatrix}$. *Find its eigenvalues.*

You need to solve

$$\begin{aligned} 0 &= \det \left(\begin{pmatrix} 1 & 2 & 4 \\ 0 & 4 & 7 \\ 0 & 0 & 6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \\ &= \det \begin{pmatrix} 1-\lambda & 2 & 4 \\ 0 & 4-\lambda & 7 \\ 0 & 0 & 6-\lambda \end{pmatrix} = (1-\lambda)(4-\lambda)(6-\lambda). \end{aligned}$$

Thus the eigenvalues are just the diagonal entries of the original matrix. You can see it would work this way with any such matrix. These matrices are called **upper triangular.** Stated precisely, a matrix A is upper triangular if $A_{ij} = 0$ for all $i > j$. Similarly, it is easy to find the eigenvalues for a lower triangular matrix, on which has all zeros above the main diagonal.

12.1.5 DEFECTIVE AND NON-DEFECTIVE MATRICES

Definition 12.1.9 *By the fundamental theorem of algebra, it is possible to write the characteristic equation in the form*

$$(\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_m)^{r_m} = 0$$

where r_j is some integer no smaller than 1. Thus the eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_m$. The ***algebraic multiplicity*** *of λ_j is defined to be r_j.*

Example 12.1.10 *Consider the matrix*

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{12.6}$$

What is the algebraic multiplicity of the eigenvalue $\lambda = 1$?

In this case the characteristic equation is

$$\det(A - \lambda I) = (1-\lambda)^3 = 0$$

or equivalently,

$$\det(\lambda I - A) = (\lambda - 1)^3 = 0.$$

Therefore, λ is of algebraic multiplicity 3.

Definition 12.1.11 *The **geometric multiplicity** of an eigenvalue is the dimension of the eigenspace,* $\ker(A - \lambda I)$.

Example 12.1.12 *Find the geometric multiplicity of* $\lambda = 1$ *for the matrix in 12.6.*

We need to solve

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix which must be row reduced to get this solution is therefore,

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This requires $z = y = 0$ and is arbitrary. Thus the eigenspace is $t\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T, t \in \mathbb{F}$. It follows the geometric multiplicity of $\lambda = 1$ is 1.

Definition 12.1.13 *An $n \times n$ matrix is called **defective** if the geometric multiplicity is not equal to the algebraic multiplicity for some eigenvalue. Sometimes such an eigenvalue for which the geometric multiplicity is not equal to the algebraic multiplicity is called a defective eigenvalue. If the geometric multiplicity for an eigenvalue equals the algebraic multiplicity, the eigenvalue is sometimes referred to as non-defective.*

Here is another more interesting example of a defective matrix.

Example 12.1.14 *Let*

$$A = \begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix}.$$

Find the eigenvectors and eigenvalues.

In this case the eigenvalues are $3, 6, 6$ where we have listed twice because it is a zero of algebraic multiplicity two, the characteristic equation being

$$(\lambda - 3)(\lambda - 6)^2 = 0.$$

It remains to find the eigenvectors for these eigenvalues. First consider the eigenvectors for $\lambda = 3$. You must solve

$$\left(\begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix} - 3\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix is

$$\left(\begin{array}{ccc|c} -1 & -2 & -1 & 0 \\ -2 & -4 & -2 & 0 \\ 14 & 25 & 11 & 0 \end{array}\right)$$

and the row reduced echelon form is

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

so the eigenvectors are nonzero vectors of the form $\begin{pmatrix} t & -t & t \end{pmatrix}^T = t\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}^T$.

Next consider the eigenvectors for $\lambda = 6$. This requires you to solve

$$\left(\begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix} - 6\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and the augmented matrix for this system of equations is

$$\begin{pmatrix} -4 & -2 & -1 & | & 0 \\ -2 & -7 & -2 & | & 0 \\ 14 & 25 & 8 & | & 0 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & \frac{1}{8} & 0 \\ 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the eigenvectors for $\lambda = 6$ are of the form $t\begin{pmatrix} -\frac{1}{8} & -\frac{1}{4} & 1 \end{pmatrix}^T$ or simply as $t\begin{pmatrix} -1 & -2 & 8 \end{pmatrix}^T$ where $t \in \mathbb{F}$.

Note that in this example the eigenspace for the eigenvalue $\lambda = 6$ is of dimension 1 because there is only one parameter. However, this eigenvalue is of multiplicity two as a root to the characteristic equation. Thus this eigenvalue is a defective eigenvalue. However, the eigenvalue 3 is nondefective. The matrix is defective because it has a defective eigenvalue.

The word, defective, seems to suggest there is something wrong with the matrix. This is in fact the case. Defective matrices are a lot of trouble in applications and we may wish they never occurred. However, they do occur as the above example shows. When you study linear systems of differential equations, you will have to deal with the case of defective matrices and you will see how awful they are. The reason these matrices are so horrible to work with is that it is impossible to obtain a basis of eigenvectors. When you study differential equations, solutions to first order systems are expressed in terms of eigenvectors of a certain matrix times $e^{\lambda t}$ where λ is an eigenvalue. In order to obtain a general solution of this sort, you must have a basis of eigenvectors. For a defective matrix, such a basis does not exist and so you have to go to something called generalized eigenvectors. Unfortunately, it is **never** explained in beginning differential equations courses why there are enough generalized eigenvectors and eigenvectors to represent the general solution. In fact, this reduces to a difficult question in linear algebra equivalent to the existence of something called the Jordan Canonical form which is much more difficult than everything discussed in the entire differential equations course. If you become interested in this, see Appendix A.

Ultimately, the algebraic issues which will occur in differential equations are a red herring anyway. The real issues relative to existence of solutions to systems of ordinary differential equations are analytical, having much more to do with calculus than with linear algebra although this will likely not be made clear either when you take a beginning differential equations class.

In terms of algebra, this lack of a basis of eigenvectors says that it is impossible to obtain a diagonal matrix which is similar to the given matrix.

Although there may be repeated roots to the characteristic equation, 12.2 and it is not known whether the matrix is defective in this case, there is an important theorem which holds when considering eigenvectors which correspond to distinct eigenvalues.

Theorem 12.1.15 *Suppose* $M\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, \cdots, r$, $\mathbf{v}_i \neq 0$, *and that if* $i \neq j$, *then* $\lambda_i \neq \lambda_j$. *Then the set of eigenvectors,* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *is linearly independent.*

Proof. Suppose the claim of the lemma is not true. Then there exists a subset of this set of vectors

$$\{\mathbf{w}_1, \cdots, \mathbf{w}_r\} \subseteq \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$$

such that

$$\sum_{j=1}^{r} c_j \mathbf{w}_j = \mathbf{0} \tag{12.7}$$

where each $c_j \neq 0$. Say $M\mathbf{w}_j = \mu_j \mathbf{w}_j$ where

$$\{\mu_1, \cdots, \mu_r\} \subseteq \{\lambda_1, \cdots, \lambda_k\},$$

the μ_j being distinct eigenvalues of M. Out of all such subsets, let this one be such that r is as small as possible. Then necessarily, $r > 1$ because otherwise, $c_1 \mathbf{w}_1 = \mathbf{0}$ which would imply $\mathbf{w}_1 = \mathbf{0}$, which is not allowed for eigenvectors.

Now apply M to both sides of 12.7.

$$\sum_{j=1}^{r} c_j \mu_j \mathbf{w}_j = \mathbf{0}. \tag{12.8}$$

Next pick $\mu_k \neq 0$ and multiply both sides of 12.7 by μ_k. Such a μ_k exists because $r > 1$. Thus

$$\sum_{j=1}^{r} c_j \mu_k \mathbf{w}_j = \mathbf{0} \tag{12.9}$$

Subtract the sum in 12.9 from the sum in 12.8 to obtain

$$\sum_{j=1}^{r} c_j (\mu_k - \mu_j) \mathbf{w}_j = \mathbf{0}$$

Now one of the constants $c_j (\mu_k - \mu_j)$ equals 0, when $j = k$. Therefore, r was not as small as possible after all. ■

Here is another proof in case you did not follow the above.

Theorem 12.1.16 *Suppose* $M\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, \cdots, r$, $\mathbf{v}_i \neq 0$, *and that if* $i \neq j$, *then* $\lambda_i \neq \lambda_j$. *Then the set of eigenvectors,* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *is linearly independent.*

Proof: Suppose the conclusion is not true. Then in the matrix

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{pmatrix}$$

not every column is a pivot column. Let the pivot columns be $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$, $k < r$. Then there exists $\mathbf{v} \in \{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$, $M\mathbf{v} = \lambda_{\mathbf{v}} \mathbf{v}$, $\mathbf{v} \notin \{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$, such that

$$\mathbf{v} = \sum_{i=1}^{k} c_i \mathbf{w}_i. \tag{12.10}$$

Then doing M to both sides yields

$$\lambda_{\mathbf{v}} \mathbf{v} = \sum_{i=1}^{k} c_i \lambda_{\mathbf{w}_i} \mathbf{w}_i \tag{12.11}$$

But also you could multiply both sides of 12.10 by $\lambda_{\mathbf{v}}$ to get

$$\lambda_{\mathbf{v}} \mathbf{v} = \sum_{i=1}^{k} c_i \lambda_{\mathbf{v}} \mathbf{w}_i.$$

And now subtracting this from 12.11 yields

$$\mathbf{0} = \sum_{i=1}^{k} c_i \left(\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i}\right) \mathbf{w}_i$$

and by independence of the $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$, this requires $c_i \left(\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i}\right) = 0$ for each i. Since the eigenvalues are distinct, $\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i} \neq 0$ and so each $c_i = 0$. But from 12.10, this requires $\mathbf{v} = \mathbf{0}$ which is impossible because $\mathbf{v}$ is an eigenvector and

Eigenvectors are never equal to zero!

Definition 12.1.17 *An* $n \times n$ *matrix is* A *called nondefective if and only if there exists a basis of eigenvectors for* $\mathbb{F}^n$.

In fact the geometric multiplicity is never larger than the algebraic multiplicity. Let A denote an $n \times n$ matrix in what follows and we assume there is an eigenvalue λ and $\mathbb{F}$ will denote the field of scalars.

Theorem 12.1.18 *Let* λ *be an eigenvalue for an* $n \times n$ *matrix. Then its geometric multiplicity is never larger than its algebraic multiplicity.*

Proof: Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ be a basis for the eigenspace corresponding to some λ. Then by Theorem 8.5.21, there is a longer list $\{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{u}_1, \cdots, \mathbf{u}_{n-r}\}$ which is a basis for $\mathbb{F}^n$. Thus the matrix $S \equiv \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r & \mathbf{u}_1 & \cdots & \mathbf{u}_{n-r} \end{pmatrix}$ is invertible. Say its inverse is

$$S^{-1} = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \mathbf{b}_1 & \cdots & \mathbf{b}_{n-r} \end{pmatrix}^T$$

where $\mathbf{a}_i^T \mathbf{v}_j = \delta_{ij}$. Then $S^{-1}AS$ must be

$$\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \mathbf{b}_1 & \cdots & \mathbf{b}_{n-r} \end{pmatrix}^T \begin{pmatrix} \lambda\mathbf{v}_1 & \cdots & \lambda\mathbf{v}_r & A\mathbf{u}_1 & \cdots & A\mathbf{u}_{n-r} \end{pmatrix}$$

and so $S^{-1}AS$ is of the form

$$\begin{pmatrix} D & M \\ 0 & N \end{pmatrix} \tag{*}$$

where D is an $r \times r$ diagonal matrix having λ down the main diagonal, M an $r \times (n-r)$, and N a $(n-r) \times (n-r)$. Now

$$\det\left(S^{-1}AS - \mu I\right) = \det\left(S^{-1}(A - \mu I)S\right) = \det\left(S^{-1}\right)\det(S)\det(A - \mu I) = \det(A - \mu I)$$

and so this matrix in $*$ has the same eigenvalues with the same multiplicities as the matrix . So consider the characteristic polynomial with variable μ of the matrix in $*$. Expanding repeatedly along the first column, one obtains the characteristic polynomial is of the form

$$q(\mu) = (\mu - \lambda)^r \det(\mu I - N) = 0$$

and so the multiplicity of λ is at least r. ■

This yields easily the following corollary which ties this theorem to Theorem 12.1.16.

Corollary12.1.19 *Let A be an $n \times n$ matrix and suppose the characteristic polynomial factors completely and that the eigenvalues are ela $\lambda_1, \cdots, \lambda_m$. If no eigenvalue is defective, then A is nondefective. If some eigenvalue is defective, then A is defective.*

Proof: Let $\{\mathbf{v}_j^i\}_{j=1}^{r_i} \equiv \beta_i$ be a basis for the eigenspace of λ_i. First I claim that is $\{\beta_1, \cdots, \beta_m\}$ linearly independent. To see this, suppose $\mathbf{w}_i \in \operatorname{span}(\beta_i)$ and $\sum_{i=1}^m \mathbf{w}_i = \mathbf{0}$. Then by Theorem 12.1.16, each $\mathbf{w}_i = \mathbf{0}$ since otherwise, you would have a nontrivial linear combination of eigenvectors associated with distinct eigenvalues which is $\mathbf{0}$. Now suppose

$$\sum_i \sum_j c_j^i \mathbf{v}_j^i = \mathbf{0}$$

Letting $\mathbf{w}_i = \sum_j c_j^i \mathbf{v}_j^i$, it follows from what was just observed that each $\mathbf{w}_i = \mathbf{0}$. Now the independence β_i of the vectors in implies that for each $i, c_j^i = 0$ for each j. Thus these vectors $\{\beta_1, \cdots, \beta_m\}$ are linearly independent as claimed. Letting $|\beta_i|$ denote the dimension of which equals the geometric multiplicity of λ_i and letting m_i denote the algebraic multiplicity of it follows that

$$\sum_i |\beta_i| \leq n = \sum_i m_i$$

If each $|\beta_i| = m_i$, then A is nondefective because $\{\beta_1, \cdots, \beta_m\}$ will then be a basis of eigenvectors. This is the case of no defective eigenvalues.

In case $|\beta_i| < m_i$ for some i, then A must be defective because if not, you would have a basis of eigenvectors and you could γ_j let be those which pertain to λ_i. Then γ_j would be an independent set of vectors and therefore, the number of vectors in γ_j denoted as $|\gamma_j|$ is no more than $|\beta_i|$. But then,

$$\sum_j |\gamma_j| \leq \sum_j |\beta_j| < \sum_j m_j = n$$

and so, you would have fewer than n vectors in this basis of eigenvectors which cannot occur. Hence A is defective. Thus if an eigenvalue is defective, the matrix is defective and if no eigenvalue is defective, then the matrix is non-defective. ■

12.1.6 DIAGONALIZATION

First of all, here is what it means for two matrices to be similar.

Definition 12.1.20 *Let A, B be two $n \times n$ matrices. Then they are* ***similar*** *if and only if there exists an invertible matrix S such that*

$$A = S^{-1}BS$$

Proposition 12.1.21 *Define for $n \times n$ matrices $A \sim B$ if A is similar to B. Then*

$$\begin{gathered} A \sim A, \\ \text{If } A \sim B \text{ then } B \sim A \\ \text{If } A \sim B \text{ and } B \sim C \text{ then } A \sim C \end{gathered}$$

Proof: It is clear that $A \sim A$ because you could just take $S = I$. If $A \sim B$, then for some S invertible,

$$A = S^{-1}BS \text{ and so } SAS^{-1} = B$$

But then

$$\left(S^{-1}\right)^{-1} AS^{-1} = B$$

which shows that $B \sim A$.

Now suppose $A \sim B$ and $B \sim C$. Then there exist invertible matrices S, T such that

$$A = S^{-1}BS, \ B = T^{-1}CT.$$

Therefore,

$$A = S^{-1}T^{-1}CTS = (TS)^{-1} C (TS)$$

showing that A is similar to C. ■

For your information, when $\sim$ satisfies the above conditions, it is called a similarity relation. Similarity relations are very significant in mathematics.

When a matrix is similar to a diagonal matrix, the matrix is said to be diagonalizable. I think this is one of the worst monstrosities for a word that I have ever seen. Nevertheless, it is commonly used in linear algebra. It turns out to be the same as non-defective which will follow easily from later material. The following is the precise definition.

Definition 22. *Let be an $n \times n$ matrix. Then A is* ***diagonalizable*** *if there exists an invertible matrix such S that*

$$S^{-1}AS = D$$

where D is a diagonal matrix. This means D has a zero as every entry except for the main diagonal. More precisely, $D_{ij} = 0$ unless $i = j$. Such matrices look like the following.

$$\begin{pmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{pmatrix}$$

where $*$ might not be zero.

The most important theorem about diagonalizability[9] is the following major result. First here is a simple observation.

Observation 12.1.23 *Let $S = \begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_n \end{pmatrix}$ where S is $n \times n$. Then here is the result of multiplying on the right by a diagonal matrix.*

$$\begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \mathbf{s}_1 & \cdots & \lambda_n \mathbf{s}_n \end{pmatrix}$$

This follows from the way we multiply matrices. The i^{th} entry of the j^{th} column of the product on the left is of the form $s_i \lambda_j$. Thus the column of the matrix on the left is just $\lambda_j \mathbf{s}_j$..

Theorem 12.1.24 *An $n \times n$ matrix is diagonalizable if and only if $\mathbb{F}^n$ has a basis of eigenvectors of A. Furthermore, you can take the matrix S described above, to be given as*

$$S = \begin{pmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_n \end{pmatrix}$$

where here the $\mathbf{s}_k$ are the eigenvectors in the basis for $\mathbb{F}^n$. If A is diagonalizable, the eigenvalues of are the diagonal entries of the diagonal matrix.

Proof: To say that A is diagonalizable, is to say that

$$S^{-1}AS = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

the λ_i being elements of $\mathbb{F}$. This is to say that for $S = \begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_n \end{pmatrix}$, $\mathbf{s}_k$ being the k^{th} column,

$$A \begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_n \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

which is equivalent, from the way we multiply matrices, that

$$\begin{pmatrix} A\mathbf{s}_1 & \cdots & A\mathbf{s}_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \mathbf{s}_1 & \cdots & \lambda_n \mathbf{s}_n \end{pmatrix}$$

which is equivalent to saying that the columns of S are eigenvectors and the diagonal matrix has the eigenvectors down the main diagonal. Since S^{-1} is invertible, these eigenvectors are a basis. Similarly, if there is a basis of eigenvectors, one can take them as the columns of S and reverse the above steps, finally concluding that is A diagonalizable. ■

Example 12.1.25 *Let* $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{pmatrix}$. *Find a matrix,* S *such that* $S^{-1}AS = D$, *a diagonal matrix.*

Solving $\det(\lambda I - A) = 0$ yields the eigenvalues are 2 and 6 with 2 an eigenvalue of multiplicity two. Solving $(2I - A)\mathbf{x} = \mathbf{0}$ to find the eigenvectors, you find that the eigenvectors are

$$a\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

where a, b are scalars. An eigenvector for $\lambda = 6$ is $\begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$. Let the matrix S be

$$S = \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

That is, the columns are the eigenvectors. Then

$$S^{-1} = \begin{pmatrix} -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{pmatrix}.$$

Then $S^{-1}AS =$

$$\begin{pmatrix} -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 1 & 4 & -1 \\ -2 & -4 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

We know the result from the above theorem, but it is nice to see it work in a specific example just the same. You may wonder if there is a need to find S^{-1}. The following is an example of a situation where this is needed. It is one of the major applications of diagonalizability.

Example 12.1.26 *Here is a matrix.* $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$ *Find* A^{50}.

Sometimes this sort of problem can be made easy by using diagonalization. In this case there are eigenvectors,

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

the first two corresponding to $\lambda = 1$ and the last corresponding to $\lambda = 2$. Then let the eigenvectors be the columns of the matrix, S. Thus

$$S = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Then also

$$S^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{pmatrix}$$

and $S^{-1}AS =$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} = D$$

Now it follows

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

Note that $\left(SDS^{-1}\right)^2 = SDS^{-1}SDS^{-1} = SD^2S^{-1}$ and $\left(SDS^{-1}\right)^3 = SDS^{-1}SDS^{-1}SDS^{-1} = SD^3S^{-1}$, etc. In general, you can see that

$$\left(SDS^{-1}\right)^n = SD^nS^{-1}$$

In other words, $A^n = SD^nS^{-1}$. Therefore,

$$A^{50} = SD^{50}S^{-1} = \begin{pmatrix} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}^{50} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{pmatrix}.$$

It is easy to raise a diagonal matrix to a power.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}^{50} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^{50} \end{pmatrix}.$$

It follows $A^{50} =$

$$\begin{pmatrix} 0 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2^{50} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 2^{50} & -1+2^{50} & 0 \\ 0 & 1 & 0 \\ 1-2^{50} & 1-2^{50} & 1 \end{pmatrix}.$$

That isn't too hard. However, this would have been horrendous if you had tried to multiply A^{50} by hand.

This technique of diagonalization is also important in solving the differential equations resulting from vibrations. Sometimes you have systems of differential equation and when you diagonalize an appropriate matrix, you "decouple" the equations. This is very nice. It makes hard problems trivial.

The above example is entirely typical. If $A = SDS^{-1}$ then $A^m = SD^mS^{-1}$ and it is easy to compute D^m. More generally, you can define functions of the matrix using power series in this way.

12.1.7 THE MATRIX EXPONENTIAL

When is diagonalizable, one can easily define what is meant by e^A. Here is how. You know

$$S^{-1}AS = D$$

where D is a diagonal matrix. You also know that if D is of the form

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \tag{12.12}$$

then

$$D^m = \begin{pmatrix} \lambda_1^m & & 0 \\ & \ddots & \\ 0 & & \lambda_n^m \end{pmatrix}$$

and that

$$A^m = SD^mS^{-1}$$

as shown above. Recall why this was.

$$A = SDS^{-1}$$

and so

$$A^m = \overbrace{SDS^{-1}SDS^{-1}SDS^{-1}\cdots SDS^{-1}}^{\text{n times}} = SD^mS^{-1}$$

Now formally write the following power series for e^A

$$e^A \equiv \sum_{k=0}^{\infty}\frac{A^k}{k!} = \sum_{k=0}^{\infty}\frac{SD^kS^{-1}}{k!} = S\sum_{k=0}^{\infty}\frac{D^k}{k!}S^{-1}$$

If D is given above in 12.12, the above sum is of the form

$$S\sum_{k=0}^{\infty}\begin{pmatrix} \frac{1}{k!}\lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \frac{1}{k!}\lambda_n^k \end{pmatrix}S^{-1} = S\begin{pmatrix} \sum_{k=0}^{\infty}\frac{1}{k!}\lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \sum_{k=0}^{\infty}\frac{1}{k!}\lambda_n^k \end{pmatrix}S^{-1}$$

$$= S\begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix}S^{-1}$$

and this last thing is the definition of what is meant by e^A.

Example 12.1.27. *Let*

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

Find e^A.

The eigenvalues happen to be $1, 2, 3$ and eigenvectors associated with these eigenvalues are

$$\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \leftrightarrow 2, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \leftrightarrow 1, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow 3$$

Then let

$$S = \begin{pmatrix} -1 & 0 & -1 \\ -1 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and so

$$S^{-1} = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Then the matrix exponential is

$$\begin{pmatrix} -1 & 0 & -1 \\ -1 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} e^2 & 0 & 0 \\ 0 & e^1 & 0 \\ 0 & 0 & e^3 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} e^2 & e^2 - e^3 & e^2 - e^3 \\ e^2 - e & e^2 & e^2 - e \\ -e^2 + e & -e^2 + e^3 & -e^2 + e + e^3 \end{pmatrix}$$

Isn't that nice? You could also talk about $\sin(A)$ or $\cos(A)$ etc. You would just have to use a different power series.

This matrix exponential is actually a useful idea when solving autonomous systems of first order linear differential equations. These are equations which are of the form

$$\mathbf{x}' = A\mathbf{x}$$

where $\mathbf{x}$ is a vector in $\mathbb{R}^n$ or $\mathbb{C}^n$ and A is an $n \times n$ matrix. Then it turns out that the solution to the above system of equations is $\mathbf{x}(t) = e^{At}\mathbf{c}$ where $\mathbf{c}$ is a constant vector.

12.1.8 COMPLEX EIGENVALUES

Sometimes you have to consider eigenvalues which are complex numbers. This occurs in differential equations for example. You do these problems exactly the same way as you do the ones in which the eigenvalues are real. Here is an example.

Example 12.1.28 *Find the eigenvalues and eigenvectors of the matrix*

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

You need to find the eigenvalues. Solve

$$\det\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0.$$

This reduces to $(\lambda - 1)\left(\lambda^2 - 4\lambda + 5\right) = 0$. The solutions are $\lambda = 1, \lambda = 2 + i, \lambda = 2 - i$.

There is nothing new about finding the eigenvectors for so consider the eigenvalue $\lambda = 2+i$. You need to solve

$$\left((2+i)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In other words, you must consider the augmented matrix

$$\left(\begin{array}{ccc|c} 1+i & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & -1 & i & 0 \end{array}\right)$$

for the solution. Divide the top row by $(1+i)$ and then take $-i$ times the second row and add to the bottom. This yields

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Now multiply the second row by $-i$ to obtain

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, the eigenvectors are of the form $t\begin{pmatrix} 0 & i & 1 \end{pmatrix}^T$. You should find the eigenvectors for $\lambda = 2 - i$. These are $t\begin{pmatrix} 0 & -i & 1 \end{pmatrix}^T$. As usual, if you want to get it right you had better check it.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1-2i \\ 2-i \end{pmatrix} = (2-i)\begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix}$$

so it worked.

12.2 SOME APPLICATIONS OF EIGENVALUES AND EIGENVECTORS

12.2.1 PRINCIPAL DIRECTIONS

Recall that $n \times n$ matrices can be considered as linear transformations. If F is a 3×3 real matrix having positive determinant, it can be shown that $F = RU$ where R is a rotation matrix and U is a symmetric real matrix having positive eigenvalues. An application of this wonderful result, known to mathematicians as the **right polar factorization**, is to continuum mechanics where a chunk of material is identified with a set of points in three dimensional space.

The linear transformation, F in this context is called the **deformation gradient** and it describes the local deformation of the material. Thus it is possible to consider this deformation in terms of two processes, one which distorts the material and the other which just rotates it. It is the matrix U which is responsible for stretching and compressing. This is why in elasticity, the stress is often taken to depend on U which is known in this context as the right **Cauchy Green strain tensor.** In this context, the eigenvalues will always be positive. The symmetry of U allows the proof of a theorem which says that if λ_M is the largest eigenvalue, then in every other direction other than the one corresponding to the eigenvector for λ_M the material is stretched less than λ_M and if λ_m is the smallest eigenvalue, then in every other direction other than the one corresponding to an eigenvector of λ_M the material is stretched more than λ_m. This process of writing a matrix as a product of two such matrices, one of which preserves distance and the other which distorts is also important in applications to geometric measure theory an interesting field of study in mathematics and to the study of quadratic forms which occur in many applications such as statistics. Here we are emphasizing the application to mechanics in which the eigenvectors of the symmetric matrix U determine the **principal directions**, those directions in which the material is stretched the most or the least.

Example 12.2.1 *Find the principal directions determined by the matrix*

$$\begin{pmatrix} \frac{29}{11} & \frac{6}{11} & \frac{6}{11} \\ \frac{6}{11} & \frac{41}{44} & \frac{19}{44} \\ \frac{6}{11} & \frac{19}{44} & \frac{41}{44} \end{pmatrix}$$

The eigenvalues are $3, 1,$ *and* $\frac{1}{2}$.

It is nice to be given the eigenvalues. The largest eigenvalue is 3 which means that in the direction determined by the eigenvector associated with 3 the stretch is three times as large. The smallest eigenvalue is $1/2$ and so in the direction determined by the eigenvector for $1/2$ the material is stretched by a factor of $1/2$, becoming locally half as long. It remains to find these directions. First consider the eigenvector for 3. It is necessary to solve

$$\left(3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{29}{11} & \frac{6}{11} & \frac{6}{11} \\ \frac{6}{11} & \frac{41}{44} & \frac{19}{44} \\ \frac{6}{11} & \frac{19}{44} & \frac{41}{44} \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus the augmented matrix for this system of equations is

$$\left(\begin{array}{ccc|c} \frac{4}{11} & -\frac{6}{11} & -\frac{6}{11} & 0 \\ -\frac{6}{11} & \frac{91}{44} & -\frac{19}{44} & 0 \\ -\frac{6}{11} & -\frac{19}{44} & \frac{91}{44} & 0 \end{array} \right)$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the principal direction for the eigenvalue, 3 in which the material is stretched to the maximum extent is

$$\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

A direction vector (or unit vector) in this direction is

$$\begin{pmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{pmatrix}.$$

You should show that the direction in which the material is compressed the most is in the direction

$$\begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Note this is meaningful information which you would have a hard time finding without the theory of eigenvectors and eigenvalues.

12.2.2 MIGRATION MATRICES

There are applications which are of great importance which feature only one eigenvalue.

Definition 12.2.2 *Let locations be denoted by the numbers $1, 2, \cdots, n$. Also suppose it is the case that each year a_{ij} denotes the proportion of residents in location j which move to location i. Also suppose no one escapes or emigrates from without these n locations. This last assumption requires $\sum_i a_{ij} = 1$. Such matrices in which the columns are nonnegative numbers which sum to one are called **Markov matrices.** In this context describing migration, they are also called **migration matrices.***

Example 12.2.3 *Here is an example of one of these matrices.*

$$\begin{pmatrix} .4 & .2 \\ .6 & .8 \end{pmatrix}$$

Thus if it is considered as a migration matrix, $.4$ is the proportion of residents in location 1 which stay in location one in a given time period while $.6$ is the proportion of residents in location 1 which move to location 2 and $.2$ is the proportion of residents in location 2 which move to location 1. Considered as a Markov matrix, these numbers are usually identified with probabilities.

If $\mathbf{v} = (x_1, \cdots, x_n)^T$ where x_i is the population of location i at a given instant, you obtain the population of location i one year later by computing $\sum_j a_{ij} x_j = (A\mathbf{v})_i$. Therefore, the population of location i after years is $\left(A^k \mathbf{v}\right)_i$. An obvious application of this would be to a situation in which you rent trailers which can go to various parts of a city and you observe through experiments the proportion of trailers which go from point i to point j in a single day. Then you might want to find how many trailers would be in all the locations after 8 days.

Proposition 12.2.4 *Let $A = (a_{ij})$ be a migration matrix. Then 1 is always an eigenvalue for A.*

Proof: Remember that $\det\left(B^T\right) = \det(B)$. Therefore,

$$\det(A - \lambda I) = \det\left((A - \lambda I)^T\right) = \det\left(A^T - \lambda I\right)$$

because $I^T = I$. Thus the characteristic equation for A is the same as the characteristic equation for A^T and so A and A^T have the same eigenvalues. We will show that is an eigenvalue for A^T and then it will follow that is an eigenvalue for A.

Remember that for a migration matrix, $\sum_i a_{ij} = 1$. Therefore, if $A^T = (b_{ij})$ so $b_{ij} = a_{ji}$, it follows that

$$\sum_j b_{ij} = \sum_j a_{ji} = 1.$$

Therefore, from matrix multiplication,

$$A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_j b_{ij} \\ \vdots \\ \sum_j b_{ij} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

which shows that $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenvector for A^T corresponding to the eigenvalue, $\lambda = 1$. As explained above, this shows that $\lambda = 1$ is an eigenvalue for A because A and A^T have the same eigenvalues. ■

Example 12.2.5 *Consider the migration matrix* $\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}$ *for locations* $1, 2,$ *and* 3. *Suppose initially there are 100 residents in location 1, 200 in location 2 and 400 in location 4. Find the population in the three locations after 10 units of time.*

From the above, it suffices to consider

$$\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}^{10} \begin{pmatrix} 100 \\ 200 \\ 400 \end{pmatrix} = \begin{pmatrix} 115.08582922 \\ 120.13067244 \\ 464.78349834 \end{pmatrix}$$

Of course you would need to round these numbers off.

A related problem asks for how many there will be in the various locations after a long time. It turns out that if some power of the migration matrix has all positive entries, then there is a limiting vector $\mathbf{x} = \lim_{k\to\infty} A^k \mathbf{x}_0$ where $\mathbf{x}_0$ is the initial vector describing the number of inhabitants in the various locations initially. This vector will be an eigenvector for the eigenvalue 1 because

$$\mathbf{x} = \lim_{k\to\infty} A^k \mathbf{x}_0 = \lim_{k\to\infty} A^{k+1} \mathbf{x}_0 = A \lim_{k\to\infty} A^k \mathbf{x} = A\mathbf{x},$$

and the sum of its entries will equal the sum of the entries of the initial vector because this sum is preserved for every multiplication by A since

$$\sum_i \sum_j a_{ij} x_j = \sum_j x_j \left(\sum_i a_{ij} \right) = \sum_j x_j.$$

Here is an example. It is the same example as the one above but here it will involve the long time limit.

Example 12.2.6 *Consider the migration matrix* $\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}$ *for locations* $1, 2,$ *and* 3. *Suppose initially there are 100 residents in location 1, 200 in location 2 and 400 in location 4. Find the population in the three locations after a long time.*

You just need to find the eigenvector which goes with the eigenvalue and then normalize it so the sum of its entries equals the sum of the entries of the initial vector. Thus you need to find a solution to

$$\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The augmented matrix is

$$\left(\begin{array}{ccc|c} .4 & 0 & -.1 & 0 \\ -.2 & .2 & 0 & 0 \\ -.2 & -.2 & .1 & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -.25 & 0 \\ 0 & 1 & -.25 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore, the eigenvectors are

$$s\begin{pmatrix} (1/4) \\ (1/4) \\ 1 \end{pmatrix}$$

and all that remains is to choose the value of such that

$$\frac{1}{4}s + \frac{1}{4}s + s = 100 + 200 + 400$$

This yields $s = \frac{1400}{3}$ and so the long time limit would equal

$$\frac{1400}{3}\begin{pmatrix} (1/4) \\ (1/4) \\ 1 \end{pmatrix} = \begin{pmatrix} 116.666\,666\,666\,666\,7 \\ 116.666\,666\,666\,666\,7 \\ 466.666\,666\,666\,666\,7 \end{pmatrix}.$$

You would of course need to round these numbers off. You see that you are not far off after just 10 units of time. Therefore, you might consider this as a useful procedure because it is probably easier to solve a simple system of equations than it is to raise a matrix to a large power.

Example 12.2.7 *Suppose a migration matrix is* $\begin{pmatrix} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{pmatrix}$. *Find the comparison between the populations in the three locations after a long time.*

This amounts to nothing more than finding the eigenvector for $\lambda = 1$. Solve

$$\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix is

$$\begin{pmatrix} \frac{4}{5} & -\frac{1}{2} & -\frac{1}{5} & | & 0 \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} & | & 0 \\ -\frac{11}{20} & -\frac{1}{4} & \frac{7}{10} & | & 0 \end{pmatrix}$$

The row echelon form is

$$\begin{pmatrix} 1 & 0 & -\frac{16}{19} & 0 \\ 0 & 1 & -\frac{18}{19} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so an eigenvector is

$$\begin{pmatrix} 16 \\ 18 \\ 19 \end{pmatrix}.$$

Thus there will be $\frac{18}{16}^{th}$ more in location 2 than in location 1. There will be $\frac{19}{18}^{th}$ more in location 3 than in location 2.

You see the eigenvalue problem makes these sorts of determinations fairly simple.

There are many other things which can be said about these sorts of **migration problems.** They include things like the gambler's ruin problem which asks for the probability that a compulsive gambler will eventually lose all his money. However those problems are not so easy although they still involve eigenvalues and eigenvectors.

12.2.3 DISCRETE DYNAMICAL SYSTEMS

The migration matrices discussed above give an example of a discrete dynamical system. They are discrete, not because they are somehow tactful and polite but because they involve discrete values taken at a sequence of points rather than on a whole interval of time. An example of a situation which can be studied in this way is a predator prey model. Consider the following model where x is the number of prey and y the number of predators. These are functions of $k \in \mathbb{N}$ where $1, 2, \cdots$ are the ends of intervals of time which may be of interest in the problem.

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} A & -B \\ C & D \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

This says that x increases if there are more x and decreases as there are more y. As for y, it increases if there are more y and also if there are more x.

Example 12.2.8 *Suppose a dynamical system is of the form*

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} 1.5 & -0.5 \\ 1.0 & 0 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

Find solutions to the dynamical system for given initial conditions.

In this case, the eigenvalues of the matrix are 1, and $.5$. The matrix is of the form

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & .5 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

and so given an initial condition

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

the solution to the dynamical system is

$$\begin{aligned} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & .5 \end{pmatrix}^n \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (.5)^n \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= \begin{pmatrix} y_0\left((.5)^n - 1\right) - x_0\left((.5)^n - 2\right) \\ y_0\left(2(.5)^n - 1\right) - x_0\left(2(.5)^n - 2\right) \end{pmatrix} \end{aligned}$$

In the limit $n \to \infty$, as you get

$$\begin{pmatrix} 2x_0 - y_0 \\ 2x_0 - y_0 \end{pmatrix}$$

Thus for large n,

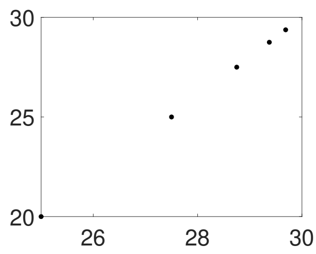
$$\begin{pmatrix} x(n) \\ y(n) \end{pmatrix} \approx \begin{pmatrix} 2x_0 - y_0 \\ 2x_0 - y_0 \end{pmatrix}$$

Letting the initial condition be

$$\begin{pmatrix} 20 \\ 10 \end{pmatrix}$$

one can graph these solutions for various values of n. Here are the solutions for values n of between 1 and 5

$$\begin{pmatrix} 25.0 \\ 20.0 \end{pmatrix}\begin{pmatrix} 27.5 \\ 25.0 \end{pmatrix}\begin{pmatrix} 28.75 \\ 27.5 \end{pmatrix}\begin{pmatrix} 29.375 \\ 28.75 \end{pmatrix}\begin{pmatrix} 29.688 \\ 29.375 \end{pmatrix}$$



Another very different kind of behavior is also observed. It is possible for the ordered pairs to spiral around the origin.

Example 12.2.9 *Suppose a dynamical system is of the form*

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} 0.7 & 0.7 \\ -0.7 & 0.7 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

Find solutions to the dynamical system for given initial conditions.

In this case, the eigenvalues are complex, $.7+.7i$ and $.7-.7i$. Suppose the initial condition is

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

what is a formula for the solutions to the dynamical system? Some computations show that the eigen pairs are

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \longleftrightarrow .7+.7i, \begin{pmatrix} 1 \\ -i \end{pmatrix} \longleftrightarrow .7-.7i$$

Thus the matrix is of the form

$$\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} .7+.7i & 0 \\ 0 & .7-.7i \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2}i \\ \frac{1}{2} & \frac{1}{2}i \end{pmatrix}$$

and so,

$$\begin{pmatrix} x(n) \\ y(n) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} (.7+.7i)^n & 0 \\ 0 & (.7-.7i)^n \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2}i \\ \frac{1}{2} & \frac{1}{2}i \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

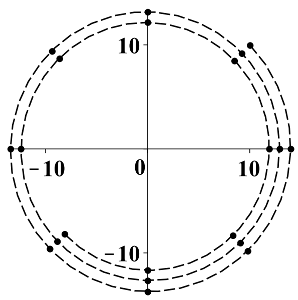
The explicit solution is given by

$$\begin{pmatrix} x_0 \left(\frac{1}{2}((0.7-0.7i))^n + \frac{1}{2}((0.7+0.7i))^n\right) + y_0 \left(\frac{1}{2}i((0.7-0.7i))^n - \frac{1}{2}i((0.7+0.7i))^n\right) \\ y_0 \left(\frac{1}{2}((0.7-0.7i))^n + \frac{1}{2}((0.7+0.7i))^n\right) - x_0 \left(\frac{1}{2}i((0.7-0.7i))^n - \frac{1}{2}i((0.7+0.7i))^n\right) \end{pmatrix}$$

Suppose the initial condition is

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \end{pmatrix}$$

Then one obtains the following sequence of values which are graphed below by letting $n = 1, 2, \cdots, 20$



In this picture, the dots are the values and the dashed line is to help to picture what is happening.

These points are getting gradually closer to the origin, but they are circling the origin in the clockwise direction as they do so. Also, since both eigenvalues are slightly smaller than 1 in absolute value,

$$\lim_{n\to\infty} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This type of behavior along with complex eigenvalues is typical of the deviations from an equilibrium point in the Lotka Volterra system of differential equations which is a famous model for predator prey interactions. These differential equations are given by

$$\begin{aligned} X' &= X(a - bY) \\ Y' &= -Y(c - dX) \end{aligned}$$

where a, b, c, d are positive constants. For example, you might have X be the population of moose and Y the population of wolves on an island.

Note how reasonable these equations are. The top says that the rate at which the moose population increases would aX be if there were no predators Y. However, this is modified by multiplying instead by $(a - bY)$ because if there are predators, these will militate against the population of moose. By definition, the wolves eat the moose and when the moose is eaten, it is not around anymore to make new moose. The more predators there are, the more pronounced is this effect. As to the predator equation, you can see that the equations predict that if there are many prey around, then the rate of growth of the predators would seem to be high. However, this is modified by the term $-cY$ because if there are many predators, there would be competition for the available food supply and this would tend to decrease Y'.

The behavior near an equilibrium point, which is a point where the right side of the differential equations equals zero, is of great interest. In this case, the equilibrium point is

$$Y = \frac{a}{b}, X = \frac{c}{d}$$

Then one defines new variables according to the formula

$$x + \frac{c}{d} = X,\ Y = y + \frac{a}{b}$$

In terms of these new variables, the differential equations become

$$\begin{aligned} x' &= \left(x + \frac{c}{d}\right)\left(a - b\left(y + \frac{a}{b}\right)\right) \\ y' &= -\left(y + \frac{a}{b}\right)\left(c - d\left(x + \frac{c}{d}\right)\right) \end{aligned}$$

Multiplying out the right sides yields

$$\begin{aligned} x' &= -bxy - b\frac{c}{d}y \\ y' &= dxy + \frac{a}{b}dx \end{aligned}$$

The interest is for x, y small and so these equations are essentially equal to

$$x' = -b\frac{c}{d}y,\ y' = \frac{a}{b}dx$$

Replace x' with the difference quotient $\frac{x(t+h)-x(t)}{h}$ where h is a small positive number and y' with a similar difference quotient. For example one could have h correspond to one day or even one hour. Thus, for h small enough, the following would seem to be a good approximation to the differential equations.

$$\begin{aligned} x(t+h) &= x(t) - hb\frac{c}{d}y \\ y(t+h) &= y(t) + h\frac{a}{b}dx \end{aligned}$$

Let $1, 2, 3, \cdots$ denote the ends of discrete intervals of time having length h chosen above. Then the above equations take the form

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} 1 & \frac{-hbc}{d} \\ \frac{had}{b} & 1 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

Note that the eigenvalues of this matrix are always complex.

You are not interested in time intervals of length h for h very small. Instead, you are interested in much longer lengths of time. Thus, replacing the time interval with mh,

$$\begin{pmatrix} x(n+m) \\ y(n+m) \end{pmatrix} = \begin{pmatrix} 1 & \frac{-hbc}{d} \\ \frac{had}{b} & 1 \end{pmatrix}^m \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

For example, if $m = 2$, you would have

$$\begin{pmatrix} x(n+2) \\ y(n+2) \end{pmatrix} = \begin{pmatrix} 1 - ach^2 & -2b\frac{c}{d}h \\ 2\frac{a}{b}dh & 1 - ach^2 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

Note that the eigenvalues of the new matrix will likely still be complex. You can also notice that the upper right corner will be negative by considering higher powers of the matrix. Thus letting $1, 2, 3, \cdots$ denote the ends of discrete intervals of time, the desired discrete dynamical system is of the form

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} A & -B \\ C & D \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$$

where A, B, C, D are positive constants and the matrix will likely have complex eigenvalues because it is a power of a matrix which has complex eigenvalues.

You can see from the above discussion that if the eigenvalues of the matrix used to define the dynamical system are less than 1 in absolute value, then the origin is stable in the sense that as $n \to \infty$, the solution converges to the origin. If either eigenvalue is larger than 1 in absolute value, then the solutions to the dynamical system will usually be unbounded, unless the initial condition is chosen very carefully. The next example exhibits the case where one eigenvalue is larger than 1 and the other is smaller than 1.

Example 12.2.10 *The Fibonacci sequence is the sequence which is defined recursively in the form*

$$x(0) = 1 = x(1), \quad x(n+2) = x(n+1) + x(n)$$

This sequence is extremely important in the study of reproducing rabbits. It can be considered as a dynamical system as follows. Let $y(n) = x(n+1)$. Then the above recurrence relation can be written as

$$\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}, \quad \begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The eigenvectors and eigenvalues of the matrix are

$$\begin{pmatrix} -\frac{1}{2}\sqrt{5}-\frac{1}{2} \\ 1 \end{pmatrix} \leftrightarrow \frac{1}{2}-\frac{1}{2}\sqrt{5}, \begin{pmatrix} \frac{1}{2}\sqrt{5}-\frac{1}{2} \\ 1 \end{pmatrix} \leftrightarrow \frac{1}{2}\sqrt{5}+\frac{1}{2}$$

You can see from a short computation that one of these is smaller than 1 in absolute value while the other is larger than 1 in absolute value.

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

Then it follows that for the given initial condition the solution to this dynamical system is of the form

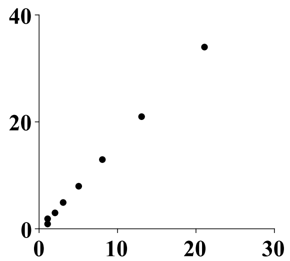
$$\begin{pmatrix} x(n) \\ y(n) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{5}-\frac{1}{2} & -\frac{1}{2}\sqrt{5}-\frac{1}{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \left(\frac{1}{2}\sqrt{5}+\frac{1}{2}\right)^n & 0 \\ 0 & \left(\frac{1}{2}-\frac{1}{2}\sqrt{5}\right)^n \end{pmatrix} \cdot$$
$$\begin{pmatrix} \frac{1}{5}\sqrt{5} & \frac{1}{10}\sqrt{5}+\frac{1}{2} \\ -\frac{1}{5}\sqrt{5} & \frac{1}{5}\sqrt{5}\left(\frac{1}{2}\sqrt{5}-\frac{1}{2}\right) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

It follows that

$$x(n) = \left(\frac{1}{2}\sqrt{5}+\frac{1}{2}\right)^n \left(\frac{1}{10}\sqrt{5}+\frac{1}{2}\right) + \left(\frac{1}{2}-\frac{1}{2}\sqrt{5}\right)^n \left(\frac{1}{2}-\frac{1}{10}\sqrt{5}\right)$$

This might not be the first thing you would think of. Here is a picture of the ordered pairs for $(x(n), y(n))$ for $n = 0, 1, \cdots, n$. There is so much more that can be said about dynamical systems.

It is a major topic of study in differential equations and what is given above is just an introduction.



12.3 THE ESTIMATION OF EIGENVALUES

There are many other important applications of eigenvalue problems. We have just given a few such applications here. As pointed out, this is a very hard problem but sometimes you don't need to find the eigenvalues exactly. There are ways to estimate the eigenvalues for matrices from just looking at the matrix. The most famous is known as **Gerschgorin's theorem.** This theorem gives a rough idea where the eigenvalues are just from looking at the matrix.

Theorem 12.3.1 *Let A be an $n \times n$ matrix. Consider the n* ***Gerschgorin discs*** *defined as*

$$D_i \equiv \left\{ \lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}.$$

Then every eigenvalue is contained in some Gerschgorin disc.

This theorem says to add up the absolute values of the entries of the i^{th} row which are off the main diagonal and form the disc centered a_{ii} at having this radius. The union of these discs contains $\sigma(A)$, the spectrum of A.

Theorem 12.3.2 *Let A be an $n \times n$ matrix. Consider the n Gerschgorin discs defined as*

$$D_i \equiv \left\{ \lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}.$$

Then every eigenvalue is contained in some Gerschgorin disc.

This theorem says to add up the absolute values of the entries of the row which are off the main diagonal and form the disc centered at having this radius. The union of these discs contains $\sigma(A)$.

Proof: Suppose $A\mathbf{x} = \lambda \mathbf{x}$ where $\mathbf{x} \neq \mathbf{0}$. Then for $A = (a_{ij})$, let $|x_k| \geq |x_j|$ for all x_j. Thus $|x_k| \neq 0$.

$$\sum_{j \neq k} a_{kj} x_j = (\lambda - a_{kk}) x_k.$$

Then

$$|x_k| \sum_{j \neq k} |a_{kj}| \geq \sum_{j \neq k} |a_{kj}| |x_j| \geq \left| \sum_{j \neq k} a_{kj} x_j \right| = |\lambda - a_{ii}| |x_k|.$$

Now dividing by $|x_k|$, it follows λ is contained in the k^{th} Gerschgorin disc. ■

Example 12.3.3 *Suppose the matrix is*

$$A = \begin{pmatrix} 21 & -16 & -6 \\ 14 & 60 & 12 \\ 7 & 8 & 38 \end{pmatrix}$$

Estimate the eigenvalues.

The exact eigenvalues are $35, 56$, and 28. The Gerschgorin disks are

$$D_1 = \{\lambda \in \mathbb{C} : |\lambda - 21| \leq 22\},$$

$$D_2 = \{\lambda \in \mathbb{C} : |\lambda - 60| \leq 26\},$$

and

$$D_3 = \{\lambda \in \mathbb{C} : |\lambda - 38| \leq 15\}.$$

Gerschgorin's theorem says these three disks contain the eigenvalues. Now 35 is in $D_3, 56$ is in D_2 and 28 is in D_1.

More can be said when the Gerschgorin disks are disjoint but this is an advanced topic which requires the theory of functions of a complex variable. If you are interested and have a background in complex variable techniques, this is in [13]

12.4 MATLAB AND EIGENVALUES

To find the eigenvalues enter A and follow with ;. Then type eig(A) and press return. It will give numerical approximation of the eigenvalues. If you want to have it find the exact values, you type eig(sym(A)) and press return. For example, if you type >>A=[1,1,0;-1,0,-1;2,1,3]; and then eig(sym(A)) and return, you will get the eigenvalues 1,1,2 listed in a column. This is correct. The matrix has a repeated eigenvalue of 1. If you want to get the eigenvectors also, you would type >>A=[1,1,0;-1,0,-1;2,1,3]; and then [V,D]=eig(sym(A)) and enter or if you want numerical answers, which will sometimes be all that is available, you would type [V,D]=eig(A). It will find the matrix V such that $AV = VD$ where D is a diagonal. In the case just considered, it will only find two columns for V because this is a defective matrix. In general, however, this would give $V^{-1}AV = D$ and the columns of V are the eigenvectors.

12.5 EXERCISES

1. State the eigenvalue problem from an algebraic perspective.
2. State the eigenvalue problem from a geometric perspective.
3. Consider the linear transformation which projects all vectors in $\mathbb{R}^2$ onto the span of the vector $(1, 2)$. Show that the matrix of this linear transformation is

   $$\begin{pmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{pmatrix}$$

   Now based on geometric considerations only, show that is an eigenvalue and that an eigenvector is $(1, 2)^T$. Also explain why 0 will also be an eigenvalue.

4. If A is the matrix of a linear transformation which rotates all vectors in $\mathbb{R}^2$ through $30°$, explain why A cannot have any real eigenvalues.
5. If A is an $n \times n$ matrix and c is a nonzero constant, compare the eigenvalues of A and cA.

6. If A is an invertible $n \times n$ matrix, compare the eigenvalues of A and A^{-1}. More generally, for m an arbitrary integer, compare the eigenvalues of A and A^m.
7. Let A, B be invertible $n \times n$ matrices which commute. That is, $AB = BA$. Suppose $\mathbf{x}$ is an eigenvector of B. Show that then must $A\mathbf{x}$ also be an eigenvector for B.
8. Suppose A is an $n \times n$ matrix and it satisfies $A^m = A$ for some m a positive integer larger than 1. Show that if λ is an eigenvalue of A then $|\lambda|$ equals either 0 or
9. Show that if $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \lambda\mathbf{y}$, then whenever are a, b scalars,

$$A\left(a\mathbf{x} + b\mathbf{y}\right) = \lambda\left(a\mathbf{x} + b\mathbf{y}\right).$$

Does this imply that $a\mathbf{x} + b\mathbf{y}$ is an eigenvector? Explain.

10. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -1 & -1 & 7 \\ -1 & 0 & 4 \\ -1 & -1 & 5 \end{pmatrix}.$$

Determine whether the matrix is defective.

11. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -3 & -7 & 19 \\ -2 & -1 & 8 \\ -2 & -3 & 10 \end{pmatrix}.$$

Determine whether the matrix is defective.

12. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -7 & -12 & 30 \\ -3 & -7 & 15 \\ -3 & -6 & 14 \end{pmatrix}.$$

Determine whether the matrix is defective.

13. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 7 & -2 & 0 \\ 8 & -1 & 0 \\ -2 & 4 & 6 \end{pmatrix}.$$

Determine whether the matrix is defective.

14. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 3 & -2 & -1 \\ 0 & 5 & 1 \\ 0 & 2 & 4 \end{pmatrix}.$$

Determine whether the matrix is defective.

15. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 6 & 8 & -23 \\ 4 & 5 & -16 \\ 3 & 4 & -12 \end{pmatrix}$$

Determine whether the matrix is defective.

16. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 5 & 2 & -5 \\ 12 & 3 & -10 \\ 12 & 4 & -11 \end{pmatrix}.$$

Determine whether the matrix is defective.

17. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 20 & 9 & -18 \\ 6 & 5 & -6 \\ 30 & 14 & -27 \end{pmatrix}.$$

Determine whether the matrix is defective.

18. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 1 & 26 & -17 \\ 4 & -4 & 4 \\ -9 & -18 & 9 \end{pmatrix}.$$

Determine whether the matrix is defective.

19. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 3 & -1 & -2 \\ 11 & 3 & -9 \\ 8 & 0 & -6 \end{pmatrix}.$$

Determine whether the matrix is defective.

20. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -2 & 1 & 2 \\ -11 & -2 & 9 \\ -8 & 0 & 7 \end{pmatrix}.$$

Determine whether the matrix is defective.

21. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 2 & 1 & -1 \\ 2 & 3 & -2 \\ 2 & 2 & -1 \end{pmatrix}.$$

Determine whether the matrix is defective.

22. Find the complex eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 4 & -2 & -2 \\ 0 & 2 & -2 \\ 2 & 0 & 2 \end{pmatrix}.$$

23. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 9 & 6 & -3 \\ 0 & 6 & 0 \\ -3 & -6 & 9 \end{pmatrix}.$$

Determine whether the matrix is defective.

24. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 4 & -2 & -2 \\ 0 & 2 & -2 \\ 2 & 0 & 2 \end{pmatrix}$.

Determine whether the matrix is defective.

25. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix} -4 & 2 & 0 \\ 2 & -4 & 0 \\ -2 & 2 & -2 \end{pmatrix}$.

Determine whether the matrix is defective.

26. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 1 & 1 & -6 \\ 7 & -5 & -6 \\ -1 & 7 & 2 \end{pmatrix}$.

Determine whether the matrix is defective.

27. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 4 & 2 & 0 \\ -2 & 4 & 0 \\ -2 & 2 & 6 \end{pmatrix}$.

Determine whether the matrix is defective.

28. Let A be a real 3×3 matrix which has a complex eigenvalue of the form $a + ib$ where $b \neq 0$. Could A be defective? Explain. Either give a proof or an example.

29. Let T be the linear transformation which reflects vectors about the x axis. Find a matrix for and then find its eigenvalues and eigenvectors.
30. Let T be the linear transformation which rotates all vectors in $\mathbb{R}^2$ counterclockwise through an angle of $\pi/2$. Find a matrix of T and then find eigenvalues and eigenvectors.
31. Let A be the 2×2 matrix of the linear transformation which rotates all vectors in $\mathbb{R}^2$ through an angle of θ. For which values of θ does A have a real eigenvalue?
32. Let T be the linear transformation which reflects all vectors in $\mathbb{R}^3$ through the xy plane. Find a matrix for T and then obtain its eigenvalues and eigenvectors.
33. Find the principal direction for stretching for the matrix

$$\begin{pmatrix} \frac{13}{9} & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} & \frac{4}{15} \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} \end{pmatrix}.$$

The eigenvalues 2 are and 1.

34. Find the principal directions for the matrix

$$\begin{pmatrix} \frac{5}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{5}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

35. Suppose the migration matrix for three locations is

$$\begin{pmatrix} .5 & 0 & .3 \\ .3 & .8 & 0 \\ .2 & .2 & .7 \end{pmatrix}.$$

Find a comparison for the populations in the three locations after a long time.

36. Suppose the migration matrix for three locations is

$$\begin{pmatrix} .1 & .1 & .3 \\ .3 & .7 & 0 \\ .6 & .2 & .7 \end{pmatrix}.$$

Find a comparison for the populations in the three locations after a long time.

37. You own a trailer rental company in a large city and you have four locations, one in the South East, one in the North East, one in the North West, and one in the South West. Denote these locations by SE,NE,NW, and SW respectively. Suppose you observe that in a typical day, $.8$ of the trailers starting in SE stay in SE, .1 of the trailers in NE go to SE, .1 of the trailers in NW end up in SE, .2 of the trailers in SW end up in SE, .1 of the trailers in SE end up in NE,.7 of the trailers in NE end up in NE,.2 of the trailers in NW end up in NE,.1 of the trailers in SW end up in NE, .1 of the trailers in SE end up in NW, .1 of the trailers in NE end up in NW, .6 of the trailers in NW end up in NW, .2 of the trailers in SW end up in NW, 0 of the trailers in SE end up in SW, .1 of the trailers in NE end up in SW, .1 of the trailers in NW end up in SW, .5 of the trailers in SW end up in SW. You begin with 20 trailers in each location. Approximately how many will you have in each location after a long time? Will any location ever run out of trailers?

38. Let A be the $n\times n$, $n>1$, matrix of the linear transformation which comes from the projection $\mathbf{v}\mapsto \operatorname{proj}_{\mathbf{w}}(\mathbf{v})$. Show that A cannot be invertible. Also show that A has an eigenvalue equal to 1 and that for λ an eigenvalue, $|\lambda|\leq 1$.

39. Let $\mathbf{v}$ be a unit vector in $\mathbb{R}^n$ and let $A = I - 2\mathbf{v}\mathbf{v}^T$. Show that A has an eigenvalue equal to -1.
40. Let M be an $n \times n$ matrix and suppose $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are n eigenvectors which form a linearly independent set. Form the matrix S by making the columns these vectors. Show that S^{-1} exists and that is a **diagonal matrix** (one having zeros everywhere except on the main diagonal) having the eigenvalues of M on the main diagonal. When this can be done the matrix is **diagonalizable.** This is presented in the text. You should write it down in your own words filling in the details without looking at the text.
41. Show that a matrix M is diagonalizable if and only if it has a basis of eigenvectors. **Hint:** The first part is done in Problem 40. It only remains to show that if the matrix can be diagonalized by some matrix S giving $D = S^{-1}MS$ for D a diagonal matrix, then it has a basis of eigenvectors. Try using the columns of the matrix S. Like the last problem, you should try to do this yourself without consulting the text. These problems are a nice review of the meaning of matrix multiplication.
42. Suppose A is an $n \times n$ matrix which is **diagonally dominant.** This means

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| .$$

Show that A^{-1} must exist.

43. Is it possible for a nonzero matrix to have only 0 as an eigenvalue?
44. Let M be an $n \times n$ matrix. Then define the adjoint of M, denoted by M^* to be the transpose of the conjugate of M. For example,

$$\begin{pmatrix} 2 & i \\ 1+i & 3 \end{pmatrix}^* = \begin{pmatrix} 2 & 1-i \\ -i & 3 \end{pmatrix}.$$

A matrix M, is self adjoint if $M^* = M$. Show the eigenvalues of a self adjoint matrix are all real. If the self adjoint matrix has all real entries, it is called symmetric.

45. Suppose A is an $n \times n$ matrix consisting entirely of real entries but $a+ib$ is a complex eigenvalue having the eigenvector $\mathbf{x} + i\mathbf{y}$. Here $\mathbf{x}$ and $\mathbf{y}$ are real vectors. Show that then $a - ib$. is also an eigenvalue with the eigenvector $\mathbf{x} - i\mathbf{y}$. **Hint:** You should remember that the conjugate of a product of complex numbers equals the product of the conjugates. Here $a+ib$ is a complex number whose conjugate equals $a+ib$.
46. Recall an $n \times n$ matrix is said to be symmetric if it has all real entries and if $A = A^T$. Show the eigenvectors and eigenvalues of a real symmetric matrix are real.

47. Recall an $n \times n$ matrix is said to be skew symmetric if it has all real entries and if $A = -A^T$. Show that any nonzero eigenvalues must be of the form ib where $i^2 = -1$. In words, the eigenvalues are either 0 or pure imaginary.

48. A discreet dynamical system is of the form

$$\mathbf{x}(k+1) = A\mathbf{x}(k), \ \mathbf{x}(0) = \mathbf{x}_0$$

where A is an $n \times n$ matrix and $\mathbf{x}(k)$ is a vector in $\mathbb{R}^n$. Show first that

$$\mathbf{x}(k) = A^k \mathbf{x}_0$$

for all $k \geq 1$. If is nondefective so that it has a basis of eigenvectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ where

$$A\mathbf{v}_j = \lambda_j \mathbf{v}_j$$

you can write the initial condition $\mathbf{x}_0$ in a unique way as a linear combination of these eigenvectors. Thus

$$\mathbf{x}_0 = \sum_{j=1}^{n} a_j \mathbf{v}_j$$

Now explain why

$$\mathbf{x}(k) = \sum_{j=1}^{n} a_j A^k \mathbf{v}_j = \sum_{j=1}^{n} a_j \lambda_j^k \mathbf{v}_j$$

which gives a formula for $\mathbf{x}(k)$, the solution of the dynamical system.

49. Suppose A is an $n \times n$ matrix and let $\mathbf{v}$ be an eigenvector such that $A\mathbf{v} = \lambda\mathbf{v}$. Also suppose the characteristic polynomial of A is

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

Explain why

$$\left(A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I\right)\mathbf{v} = \mathbf{0}$$

If A is nondefective, give a very easy proof of the Cayley Hamilton theorem based on this. Recall this theorem says A satisfies its characteristic equation,

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0.$$

50. Suppose $n \times n$ an nondefective matrix A has only 1 and -1 as eigenvalues. Find A^{12}.
51. Suppose $n \times n$ the characteristic polynomial of an $n \times n$ matrix is $1 - \lambda^n$. Find A^{mn} where m is an integer. **Hint:** Note first that A is nondefective. Why?
52. Sometimes sequences come in terms of a recursion formula. An example is the Fibonacci sequence.

$$x_0 = 1 = x_1,\ x_{n+1} = x_n + x_{n-1}$$

Show this can be considered as a discreet dynamical system as follows.

$$\begin{pmatrix} x_{n+1} \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix}, \begin{pmatrix} x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Now use the technique of Problem 48 to find a formula for x_n. This was done in the chapter. Next change the initial conditions $x_0 = 0, x_1 = 1$ to and find the solution.

53. Let A be an $n \times n$ matrix having characteristic polynomial

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

Show that $a_0 = (-1)^n \det(A)$.

54. Find $\begin{pmatrix} \frac{3}{2} & 1 \\ -\frac{1}{2} & 0 \end{pmatrix}^{35}$. Next find

$$\lim_{n\to\infty} \begin{pmatrix} \frac{3}{2} & 1 \\ -\frac{1}{2} & 0 \end{pmatrix}^{n}$$

55. Find e^A where A is the matrix $\begin{pmatrix} \frac{3}{2} & 1 \\ -\frac{1}{2} & 0 \end{pmatrix}$ in the above problem.

56. Consider the dynamical system $\begin{pmatrix} x(n+1) \\ y(n+1) \end{pmatrix} = \begin{pmatrix} .8 & .8 \\ -.8 & .8 \end{pmatrix} \begin{pmatrix} x(n) \\ y(n) \end{pmatrix}$ Show the eigenvalues and eigenvectors are $0.8+0.8i \longleftrightarrow \begin{pmatrix} -i \\ 1 \end{pmatrix}, 0.8-0.8i \longleftrightarrow \begin{pmatrix} i \\ 1 \end{pmatrix}$. Find a formula for the solution to the dynamical system for given initial condition $(x_0, y_0)^T$. Show that the magnitude of $(x(n), y(n))^T$ must diverge provided the initial condition is not zero. Next graph the vector field for

$$\begin{pmatrix} .8 & .8 \\ -.8 & .8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}$$

Note that this vector field seems to indicate a conclusion different than what you just obtained. Therefore, in this context of discreet dynamical systems the consideration of such a picture is not all that reliable.

13 MATRICES AND THE INNER PRODUCT

13.1 SYMMETRIC AND ORTHOGONAL MATRICES

13.1.1 ORTHOGONAL MATRICES

Remember that to find the inverse of a matrix was often a long process. However, it was very easy to take the transpose of a matrix. For some matrices, the transpose equals the inverse and when the matrix has all real entries, and this is true, it is called an orthogonal matrix. Recall the following definition given earlier.

Definition 13.1.1 *A real* $n \times n$ *matrix* U *is called an Orthogonal matrix if* $UU^T = U^TU = I$.

Example 13.1.2 *Show the matrix*

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

is orthogonal.

$$UU^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example 13.1.3 *Let* $U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}$. *Is* U *orthogonal?*

The answer is yes. This is because the columns form an orthonormal set of vectors as well as the rows. As discussed above this is equivalent to $U^TU = I$.

$$U^TU = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

When you say that U is orthogonal, you are saying that

$$\sum_j U_{ij} U_{jk}^T = \sum_j U_{ij} U_{kj} = \delta_{ik}.$$

In words, the dot product of the i^{th} row of U with the k^{th} row gives 1 if $i = k$ and 0 if $i \neq k$. The same is true of the columns because $U^TU = I$ also. Therefore,

$$\sum_j U_{ij} U_{jk}^T = \sum_j U_{ij} U_{kj} = \delta_{ik}.$$

which says that the one column dotted with another column gives 1 if the two columns are the same and 0 if the two columns are different.

More succinctly, this states that if $\mathbf{u}_1, \cdots, \mathbf{u}_n$ are the columns of U an orthogonal matrix, then

$$\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}. \tag{13.1}$$

Definition 13.1.4 *A set of vectors,* $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ *is said to be an* ***orthonormal*** *set if 13.1*

Theorem 13.1.5 *If* $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ *is an orthonormal set of vectors then it is linearly independent.*

Proof: Using the properties of the dot product,

$$\mathbf{0} \cdot \mathbf{u} = (\mathbf{0} + \mathbf{0}) \cdot \mathbf{u} = \mathbf{0} \cdot \mathbf{u} + \mathbf{0} \cdot \mathbf{u}$$

and so, subtracting $\mathbf{0} \cdot \mathbf{u}$ from both sides yields $\mathbf{0} \cdot \mathbf{u} = 0$. Now suppose $\sum_j c_j \mathbf{u}_j = \mathbf{0}$. Then from the properties of the dot product,

$$c_k = \sum_j c_j \delta_{jk} = \sum_j c_j (\mathbf{u}_j \cdot \mathbf{u}_k) = \left(\sum_j c_j \mathbf{u}_j \right) \cdot \mathbf{u}_k = \mathbf{0} \cdot \mathbf{u}_k = 0.$$

Since k was arbitrary, this shows that each $c_k = 0$ and this has shown that if $\sum_j c_j \mathbf{u}_j = \mathbf{0}$, then each $c_j = 0$. This is what it means for the set of vectors to be linearly independent. ■

Example 13.1.6. *Let* $U = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{\sqrt{6}}{3} \end{pmatrix}$. *Is* U *an orthogonal matrix?*

The answer is yes. This is because the columns (rows) form an orthonormal set of vectors.

The importance of orthogonal matrices is that they change components of vectors relative to different Cartesian coordinate systems. Geometrically, the orthogonal matrices are exactly those which preserve all distances in the sense that if $\mathbf{x} \in \mathbb{R}^n$ and U is orthogonal, then $||U\mathbf{x}|| = ||\mathbf{x}||$ because

$$||U\mathbf{x}||^2 = (U\mathbf{x})^T U\mathbf{x} = \mathbf{x}^T U^T U\mathbf{x} = \mathbf{x}^T I\mathbf{x} = ||\mathbf{x}||^2.$$

Observation 13.1.7 *Suppose* U *is an orthogonal matrix. Then* $\det(U) = \pm 1$.

This is easy to see from the properties of determinants. Thus

$$\det(U)^2 = \det\left(U^T\right)\det(U) = \det\left(U^T U\right) = \det(I) = 1.$$

Orthogonal matrices are divided into two classes, proper and improper. The proper orthogonal matrices are those whose determinant equals 1 and the improper ones are those whose determinants equal -1. The reason for the distinction is that the improper orthogonal matrices are sometimes considered to have no physical significance since they cause a change in orientation which would correspond to material passing through itself in a non-physical manner. Thus in considering which coordinate systems must be considered in certain applications, you only need to consider those which are related by a proper orthogonal transformation. Geometrically, the linear transformations determined by the proper orthogonal matrices correspond to the composition of rotations.

13.1.2 SYMMETRIC AND SKEW SYMMETRIC MATRICES

Definition 13.1.8 *A real $n \times n$ matrix A, is symmetric if $A^T = A$. If $A = -A^T$, If then A is called* ***skew symmetric.***

Theorem 13.1.9 *The eigenvalues of a real symmetric matrix are real. The eigenvalues of a real skew symmetric matrix are 0 or pure imaginary.*

Proof: The proof of this theorem is in [13]. It is best understood as a special case of more general considerations. However, here is a proof in this special case.

Recall that for a complex number $a + ib$, the complex conjugate, denoted by $\overline{a + ib}$ is given by the formula $\overline{a + ib} = a - ib$. The notation, $\overline{\mathbf{x}}$ will denote the vector which has every entry replaced by its complex conjugate.

Suppose A is a real symmetric matrix and $A\mathbf{x} = \lambda\mathbf{x}$. Then

$$\overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x} = \left(\overline{A\mathbf{x}}\right)^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x} = \lambda\overline{\mathbf{x}}^T\mathbf{x}.$$

Dividing by $\overline{\mathbf{x}}^T\mathbf{x}$ on both sides yields $\overline{\lambda} = \lambda$ which says λ is real. (Why?)

Next suppose $A = -A^T$ so A is skew symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then

$$\overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x} = \left(\overline{A\mathbf{x}}\right)^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = -\overline{\mathbf{x}}^T A\mathbf{x} = -\lambda\overline{\mathbf{x}}^T\mathbf{x}$$

and so, dividing by $\overline{\mathbf{x}}^T\mathbf{x}$ as before, $\overline{\lambda} = -\lambda$. Letting $\lambda = a + ib$, this means $\overline{a + ib} = a - ib$. and so $a = 0$. Thus λ is pure imaginary. ■

Example 13.1.10 *Let* $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. *This is a skew symmetric matrix. Find its eigenvalues.*

Its eigenvalues are obtained by solving the equation $\det\begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1 = 0$. You see the eigenvalues are $\pm i$, pure imaginary.

Example 13.1.11 *Let* $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. *This is a symmetric matrix. Find its eigenvalues.*

Its eigenvalues are obtained by solving the equation, $\det\begin{pmatrix} 1-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} = -1 - 4\lambda + \lambda^2 = 0$ and the solution is $\lambda = 2 + \sqrt{5}$ and $\lambda = 2 - \sqrt{5}$.

Definition 13.1.12 *An* $n \times n$ *matrix* $A = (a_{ij})$ *is called a* ***diagonal matrix*** *if* $a_{ij} = 0$ *whenever* $i \neq j$. *For example, a diagonal matrix is of the form indicated below where* $*$ *denotes a number.*

$$\begin{pmatrix} * & & 0 \\ & \ddots & \\ 0 & & * \end{pmatrix}$$

Theorem 13.1.13 *Let* A *be a real symmetric matrix. Then there exists an orthogonal matrix* U *such that* $U^T AU$ *is a diagonal matrix. Moreover, the diagonal entries are the eigenvalues of* A.

Proof: The proof is given later.

Corollary 13.1.14 *If* A *is a real* $n \times n$ *symmetric matrix, then there exists an orthonormal set of eigenvectors,* $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$.

Proof: Since A is symmetric, then by Theorem 13.1.13, there exists an orthogonal matrix U such that $U^T AU = D$, a diagonal msatrix whose diagonal entries are the eigenvalues of A . Therefore, since A is symmetric and all the matrices are real,

$$\overline{D} = \overline{D^T} = \overline{U^T A^T U} = U^T A^T U = U^T AU = D$$

showing D is real because each entry of D equals its complex conjugate.[10]

Finally, let

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

where the $\mathbf{u}_i$ denote the columns of U and

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

The equation, $U^T AU = D$ implies

$$\begin{aligned} AU &= \begin{pmatrix} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{pmatrix} \\ &= UD = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix} \end{aligned}$$

where the entries denote the columns of AU and UD respectively. Therefore, $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$ and since the matrix is orthogonal, the ij^{th} entry of $U^T U$ equals δ_{ij} and so

$$\delta_{ij} = \mathbf{u}_i^T\mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j.$$

This proves the corollary because it shows the vectors $\{\mathbf{u}_i\}$ form an orthonormal basis. ■

Example 13.1.15 *Find the eigenvalues and an orthonormal basis of eigenvectors for the matrix*

$$\begin{pmatrix} \frac{19}{9} & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} & -\frac{16}{15} \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} \end{pmatrix}$$

given that the eigenvalues are 3, -1, *and 2.*

The augmented matrix which needs to be row reduced to find the eigenvectors for $\lambda = 3$ is

$$\left(\begin{array}{ccc|c} \frac{19}{9} - 3 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} - 3 & -\frac{16}{15} & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} - 3 & 0 \end{array}\right)$$

and the row reduced echelon form for this is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{2}\sqrt{5} & | & 0 \\ 0 & 1 & \frac{3}{4} & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

Therefore, eigenvectors for $\lambda = 3$ are $z\begin{pmatrix} \frac{1}{2}\sqrt{5} & -\frac{3}{4} & 1 \end{pmatrix}^T$ where $z \neq 0$.

The augmented matrix, which must be row reduced to find the eigenvectors for $\lambda = -1$, is

$$\begin{pmatrix} \frac{19}{9}+1 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & | & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5}+1 & -\frac{16}{15} & | & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45}+1 & | & 0 \end{pmatrix}$$

and the row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{2}\sqrt{5} & | & 0 \\ 0 & 1 & -3 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

Therefore, the eigenvectors for $\lambda = -1$ are $z\begin{pmatrix} \frac{1}{2}\sqrt{5} & 3 & 1 \end{pmatrix}^T$, $z \neq 0$

The augmented matrix which must be row reduced to find the eigenvectors for $\lambda = 2$ is

$$\begin{pmatrix} \frac{19}{9} - 2 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & | & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} - 2 & -\frac{16}{15} & | & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} - 2 & | & 0 \end{pmatrix}$$

and its row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & \frac{2}{5}\sqrt{5} & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

so the eigenvectors for $\lambda = 2$ are $z\begin{pmatrix} -\frac{2}{5}\sqrt{5} & 0 & 1 \end{pmatrix}^T$, $z \neq 0$.

It remains to find an orthonormal basis. You can check that the dot product of any of these vectors with another of them gives zero and so it suffices choose z in each case such that the resulting vector has length 1. First consider the vectors for $\lambda = 3$ It is required to choose z such that $z\begin{pmatrix} \frac{1}{2}\sqrt{5} & -\frac{3}{4} & 1 \end{pmatrix}^T$ is a unit vector. In other words, you need

$$z\begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix} \cdot z\begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix} = 1.$$

But the above dot product equals $\frac{45}{16}z^2$ and this equals 1 when $z = \frac{4}{15}\sqrt{5}$. Therefore, the eigenvector which is desired is $\begin{pmatrix} \frac{2}{3} & -\frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix}^T$.

Next find the eigenvector for $\lambda = -1$ The same process requires that $1 = \frac{45}{4}z^2$ which happens when $z = \frac{2}{15}\sqrt{5}$. Therefore, an eigenvector for $\lambda = -1$ which has unit length is

$$\frac{2}{15}\sqrt{5}\begin{pmatrix} \frac{1}{2}\sqrt{5} \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ \frac{2}{15}\sqrt{5} \end{pmatrix}.$$

Finally, consider $\lambda = 2$. This time you need $1 = \frac{9}{5}z^2$ which occurs when $z = \frac{1}{3}\sqrt{5}$. Therefore, the eigenvector is

$$\frac{1}{3}\sqrt{5}\begin{pmatrix} -\frac{2}{5}\sqrt{5} \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

Now recall that the vectors form an orthonormal set of vectors if the matrix having them as columns is orthogonal. That matrix is

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ \frac{4}{15}\sqrt{5} & \frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

Is this orthogonal? To find out, multiply by its transpose. Thus

$$\begin{pmatrix} \frac{2}{3} & -\frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \\ \frac{1}{3} & \frac{2}{5}\sqrt{5} & \frac{2}{15}\sqrt{5} \\ -\frac{2}{3} & 0 & \frac{1}{3}\sqrt{5} \end{pmatrix}\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ \frac{4}{15}\sqrt{5} & \frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the identity was obtained this shows the above matrix is orthogonal and that therefore, the columns form an orthonormal set of vectors. The problem asks for you to find an orthonormal basis. However, you will show in Problem 23 that an orthonormal set of n vectors in $\mathbb{R}^n$ is always a basis. Therefore, since there are three of these vectors, they must constitute a basis.

Example 13.1.16 *Find an orthonormal set of three eigenvectors for the matrix*

$$\begin{pmatrix} \frac{13}{9} & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} & \frac{4}{15} \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} \end{pmatrix}$$

given the eigenvalues are 2, *and* 1.

The eigenvectors which go with $\lambda = 2$ are obtained from row reducing the matrix

$$\left(\begin{array}{ccccc} \frac{13}{9} - 2 & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} & | & 0 \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} - 2 & \frac{4}{15} & | & 0 \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} - 2 & | & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\left(\begin{array}{ccccc} 1 & 0 & -\frac{1}{2}\sqrt{5} & | & 0 \\ 0 & 1 & -\frac{3}{4} & | & 0 \\ 0 & 0 & 0 & | & 0 \end{array}\right)$$

which shows the eigenvectors for $\lambda = 2$ are $z \left(\begin{array}{ccc} \frac{1}{2}\sqrt{5} & \frac{3}{4} & 1 \end{array} \right)^T$ and a choice for z which will produce a unit vector is $z = \frac{4}{15}\sqrt{5}$. Therefore, the vector we want is $\left(\begin{array}{ccc} \frac{2}{3} & \frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{array} \right)^T$.

Next consider the eigenvectors for $\lambda = 1$. The matrix which must be row reduced is

$$\left(\begin{array}{ccc|c} \frac{13}{9}-1 & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} & 0 \\ \frac{2}{15}\sqrt{5} & \frac{6}{5}-1 & \frac{4}{15} & 0 \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45}-1 & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & \frac{3}{10}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors are of the form $\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z & y & z \end{pmatrix}^T, y, z$ arbitrary. This is a two dimensional eigenspace.

Before going further, we want to point out that no matter how we choose y and z the resulting vector will be orthogonal to the eigenvector for $\lambda = 2$. This is a special case of a general result which states that eigenvectors for distinct eigenvalues of a symmetric matrix are orthogonal. This is explained in Problem 15. For this case you need to show the following dot product equals zero.

$$\begin{pmatrix} \frac{2}{3} \\ \frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix} \cdot \begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z \\ y \\ z \end{pmatrix} \tag{13.2}$$

This is left for you to do.

Continuing with the task of finding an orthonormal basis, Let $y = 0$ first. This results in eigenvectors of the form $\begin{pmatrix} -\frac{2}{5}\sqrt{5}z & 0 & z \end{pmatrix}^T$ and letting $z = \frac{1}{3}\sqrt{5}$ you obtain a unit vector. Thus the second vector will be

$$\begin{pmatrix} -\frac{2}{5}\sqrt{5}\left(\frac{1}{3}\sqrt{5}\right) \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

It remains to find the third vector in the orthonormal basis. This merely involves choosing y and z in 13.2 in such a way that the resulting vector has dot product with the two given vectors equal to zero. Thus you need

$$\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix} = \frac{1}{5}\sqrt{5}y + \frac{3}{5}\sqrt{5}z = 0.$$

The dot product with the eigenvector $\lambda = 2$ for is automatically equal to zero and so all that you need is the above equation. This is satisfied when $z = -\frac{1}{3}y$. Therefore, the vector we want is of the form

$$\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}\left(-\frac{1}{3}y\right) \\ y \\ \left(-\frac{1}{3}y\right) \end{pmatrix} = \begin{pmatrix} -\frac{1}{6}\sqrt{5}y \\ y \\ -\frac{1}{3}y \end{pmatrix}$$

and it only remains to choose y in such a way that this vector has unit length. This occurs when $y = \frac{2}{5}\sqrt{5}$. Therefore, the vector we want is

$$\frac{2}{5}\sqrt{5}\begin{pmatrix} -\frac{1}{6}\sqrt{5} \\ 1 \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} \end{pmatrix}.$$

The three eigenvectors which constitute an orthonormal basis are

$$\begin{pmatrix} -\frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} \end{pmatrix}, \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}, \text{ and } \begin{pmatrix} \frac{2}{3} \\ \frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

To check the work and see if this is really an orthonormal set of vectors, make them the columns of a matrix and see if the resulting matrix is orthogonal. The matrix is

$$\begin{pmatrix} -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{5}\sqrt{5} & 0 & \frac{1}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

This matrix times its transpose equals

$$\begin{pmatrix} -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{5}\sqrt{5} & 0 & \frac{1}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix} \begin{pmatrix} -\frac{1}{3} & \frac{2}{5}\sqrt{5} & -\frac{2}{15}\sqrt{5} \\ -\frac{2}{3} & 0 & \frac{1}{3}\sqrt{5} \\ \frac{2}{3} & \frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so this is indeed an orthonormal basis.

Because of the repeated eigenvalue, there would have been many other orthonormal bases which could have been obtained. It was pretty arbitrary for to take $y = 0$ in the above argument. We could just as easily have taken $z = 0$ or even $y = z = 1$. Any such change would have resulted in a different orthonormal basis. Geometrically, what is happening is the eigenspace for $\lambda = 1$. was two dimensional. It can be visualized as a plane in three dimensional space which passes through the origin. There are infinitely many different pairs of perpendicular unit vectors in this plane.

13.1.3 DIAGONALIZING A SYMMETRIC MATRIX

Recall the following definition:

Definition 13.1.17 *An $n \times n$ matrix $A = (a_{ij})$ is called a diagonal matrix if $a_{ij} = 0$ whenever $i \neq j$. For example, a diagonal matrix is of the form indicated below where $*$ denotes a number.*

$$\begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & * \end{pmatrix}$$

Definition 13.1.18 *An $n \times n$ matrix A is said to be **non defective** or **diagonalizable** if* there *exists an invertible matrix S such that $S^{-1}AS = D$ where D is a diagonal matrix as described above.*

Some matrices are non defective and some are not. As indicated in Theorem 13.1.13 if A is a real symmetric matrix, there exists an orthogonal matrix U such that $U^T AU = D$ a diagonal matrix. Therefore, every symmetric matrix is non defective because if U is an orthogonal matrix, its inverse is U^T. In the following example, this orthogonal matrix will be found.

Example 13.1.19 *Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}$. Find an orthogonal matrix U such that $U^T AU$ is a diagonal matrix.*

In this case, a tedious computation shows the eigenvalues are 2 and 1. First we will find an eigenvector for the eigenvalue 2. This involves row reducing the following augmented matrix.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 2-\frac{3}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 2-\frac{3}{2} & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so an eigenvector is $\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}^T$. However, it is desired that the eigenvectors obtained all be unit vectors and so dividing this vector by its length gives $\begin{pmatrix} 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix}^T$. Next consider the case of the eigenvalue, 1.. The matrix which needs to be row reduced in this case is

$$\left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 1-\frac{3}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1-\frac{3}{2} & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors are of the form $\begin{pmatrix} s & -t & t \end{pmatrix}^T$. Two of these which are orthonormal are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

An orthogonal matrix which works in the process is then obtained by letting these vectors be the columns.

$$\begin{pmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}.$$

It remains to verify this works. $U^T AU$ is of the form

$$\begin{pmatrix} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

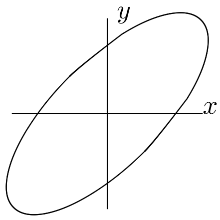
the desired diagonal matrix.

One of the applications for this technique has to do with rotation of axes so that with respect to the new axes, the graph the level curve of a quadratic form is oriented parallel to the coordinate axes. This makes it much easier to understand. This is discussed more in the exercises. However, here is a simple example.

Example 20. *Consider the following level curve.*

$$5x^2 - 6xy + 5y^2 = 8$$

Its graph is given in the following picture.



You can write this in terms of a symmetric matrix as follows.

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 8$$

Change the variables as follows.

$$\begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix} = \begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}$$

and so

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 8$$

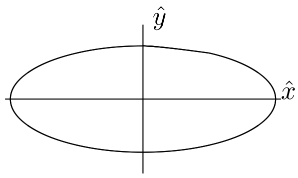
Let

$$\begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}$$

Then in terms of these new variables, you get

$$2\hat{x}^2 + 8\hat{y}^2 = 8$$

This is an ellipse which is parallel to the coordinate axes. Its graph is of the form



Thus this change of variables chooses new axes such that with respect to these new axes, the ellipse is oriented parallel to the coordinate axes. These new axes are called the principal axes.

In general a quadratic form is an expression of the form

$$\mathbf{x}^T A\mathbf{x}$$

where A is a symmetric matrix. When you write something like

$$\mathbf{x}^T A\mathbf{x} = c$$

you are considering a level surface or level curve of some sort. By diagonalizing the matrix as shown above, you can choose new variables such that in the new variables, there are no "mixed" terms like xy or $yz.$ Geometrically this has the effect of choosing new coordinate axes such that with respect to these new axes, the various axes of symmetry of the level surfaces or curves are parallel to the coordinate axes. Therefore, this is a desirable simplification. Other quadratic forms in two variables lead to parabolas or hyperbolas. In three dimensions there are also names associated with these quadratic surfaces usually involving the semi word "oid". They are typically discussed in calculus courses where they are invariably oriented parallel to the coordinate axes. However, the process of diagonalization just explained will allow one to start with one which is not oriented this way and reduce it to one which is.

13.2 FUNDAMENTAL THEORY AND GENERALIZATIONS

13.2.1 BLOCK MULTIPLICATION OF MATRICES

Consider the following problem

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

You know how to do this. You get

$$\begin{pmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{pmatrix}.$$

Now what if instead of numbers, the entries, A, B, C, D, E, F, G are matrices of a size such that the multiplications and additions needed in the above formula all make sense. Would the formula be true in this case? I will show below that this is true.

Suppose A is a matrix of the form

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{r1} & \cdots & A_{rm} \end{pmatrix} \tag{13.136}$$

where A_{ij} is a $s_i \times p_j$ matrix where s_i is constant for $j = 1, \cdots, m$ for each $i = 1, \cdots, r$. Such a matrix is called a **block matrix**, also a **partitioned matrix**. How do you get the block A_{ij}? Here is how for A an $m \times n$ matrix:

$$\overbrace{\begin{pmatrix} \mathbf{0} & I_{s_i \times s_i} & \mathbf{0} \end{pmatrix}}^{s_i \times m} A \overbrace{\begin{pmatrix} \mathbf{0} \\ I_{p_j \times p_j} \\ \mathbf{0} \end{pmatrix}}^{n \times p_j}. \tag{13.4}$$

In the block column matrix on the right, you need to have $c_j - 1$ rows of zeros above the small $p_j \times p_j$ identity matrix where the columns of A involved in A_{ij} are $c_j, \cdots, c_j + p_j - 1$ and in the block row matrix on the left, you need to have $r_i - 1$ columns of zeros to the left of the $s_i \times s_i$ identity matrix where the rows of A involved in A_{ij} are $r_i, \cdots, r_i + s_i$. An important observation to make is that the matrix on the right specifies columns to use in the block and the one on the left specifies the rows used. Thus the block A_{ij} in this case is a matrix of size $s_i \times p_j$. There is no overlap between the blocks of A. Thus the identity identity $n \times n$ matrix corresponding to multiplication on the right of A is of the form

$$\begin{pmatrix} I_{p_1 \times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_m \times p_m} \end{pmatrix}$$

these little identity matrices don't overlap. A similar conclusion follows from consideration of the matrices $I_{s_i \times s_i}$.

Next consider the question of multiplication of two block matrices. Let B be a block matrix of the form

$$\begin{pmatrix} B_{11} & \cdots & B_{1p} \\ \vdots & \ddots & \vdots \\ B_{r1} & \cdots & B_{rp} \end{pmatrix} \tag{13.5}$$

and A is a block matrix of the form

$$\begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{p1} & \cdots & A_{pm} \end{pmatrix} \tag{13.6}$$

and that for all i, j, it makes sense to multiply $B_{is}A_{sj}$ for all $s \in \{1, \cdots, p\}$. (That is the two matrices, B_{is} and A_{sj} are conformable.) and that for fixed ij, it follows $B_{is}A_{sj}$ is the same size for each s so that it makes sense to write $\sum_s B_{is}A_{sj}$.

The following theorem says essentially that when you take the product of two matrices, you can do it two ways. One way is to simply multiply them forming BA. The other way is to partition both matrices, formally multiply the blocks to get another block matrix and this one will be BA partitioned. Before presenting this theorem, here is a simple lemma which is really a special case of the theorem.

Lemma 13.2.1 *Consider the following product.*

$$\begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I & \mathbf{0} \end{pmatrix}$$

where the first is $n \times r$ and the second is $r \times n$. The small identity matrix I is an $r \times r$ matrix and there are l zero rows above I and l zero columns to the left of I in the right matrix. Then the product of these matrices is a block matrix of the form

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Proof: From the definition of the way you multiply matrices, the product is

$$\left(\begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \cdots \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_1 \cdots \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_r \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \cdots \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \right)$$

which yields the claimed result. In the formula $\mathbf{e}_j$ refers to the column vector of length r which has a 1 in the j^{th} position. ■

Theorem 13.2.2 *Let B be a $q \times p$ block matrix as in 13.5 and let A be a $p \times n$ block matrix as in 13.6 such that B_{is} is conformable with A_{sj} and each product, $B_{is}A_{sj}$ for $s = 1, \cdots, p$ is of the same size so they can be added. Then BA. can be obtained as a block matrix such that the ij^{th} block is of the form*

$$\sum_s B_{is} A_{sj}. \tag{13.7}$$

Proof: From 13.4

$$B_{is}A_{sj} = \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

where here it is assumed B_{is} is $r_i \times p_s$ and A_{sj} is $p_s \times q_j$. The product involves the s^{th} block in the i^{th} row of blocks for B and the s^{th} block in the j^{th} column of A. Thus there are the same number of rows above the $I_{p_s \times p_s}$ as there are columns to the left of $I_{p_s \times p_s}$ in those two inside matrices. Then from Lemma 13.2.1

$$\begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Since the blocks of small identity matrices do not overlap,

$$\sum_s \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} I_{p_1 \times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_p \times p_p} \end{pmatrix} = I$$

and so $\sum_s B_{is} A_{sj} =$

$$\sum_s \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \sum_s \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} BIA \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} BA \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

which equals the ij^{th} block of BA. Hence the ij^{th} block of BA equals the formal multiplication according to matrix multiplication, $\sum_s B_{is} A_{sj}$. ■

Example 13.2.3 *Let an $n \times n$ matrix have the form*

$$A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix}$$

where P is $n-1 \times n-1$. Multiply it by

$$B = \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix}$$

where B is also an $n \times n$ matrix and Q is $n-1 \times n-1$.

You use block multiplication

$$\begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix} \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix} = \begin{pmatrix} ap + \mathbf{br} & a\mathbf{q} + \mathbf{b}Q \\ p\mathbf{c} + P\mathbf{r} & \mathbf{cq} + PQ \end{pmatrix}$$

Note that this all makes sense. For example, $\mathbf{b} = 1 \times n - 1$ and $\mathbf{r} = n - 1 \times 1$ so $\mathbf{br}$ is a 1×1. Similar considerations apply to the other blocks.

Here is an interesting and significant application of block multiplication. In this theorem, $p_M(t)$ denotes the characteristic polynomial, $\det(tI - M)$. Thus the zeros of this polynomial are the eigenvalues of the matrix M.

Theorem 13.2.4 *Let A be an $m \times n$ matrix and let be B an $n \times m$ matrix for $m \leq n$. Then*

$$p_{BA}(t) = t^{n-m} p_{AB}(t),$$

so the eigenvalues of BA and AB are the same including multiplicities except that BA has $n - m$ extra zero eigenvalues.

Proof: Use block multiplication to write

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}$$

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

Since the two matrices above are similar it follows that $\begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$ and $\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$ have the same characteristic polynomials. Therefore, noting that BA is an $n \times n$ matrix and AB is an $m \times m$ matrix,

$$t^m \det(tI - BA) = t^n \det(tI - AB)$$

and so $\det(tI - BA) = p_{BA}(t) = t^{n-m} \det(tI - AB) = t^{n-m} p_{AB}(t)$.

13.2.2 ORTHONORMAL BASES, GRAM SCHMIDT PROCESS

Not all bases for $\mathbb{F}^n$ are created equal. Recall $\mathbb{F}$ equals either $\mathbb{C}$ or $\mathbb{R}$ and the dot product is given by

$$\mathbf{x} \cdot \mathbf{y} \equiv (\mathbf{x}, \mathbf{y}) \equiv \langle \mathbf{x}, \mathbf{y} \rangle = \sum_j x_j \overline{y_j}.$$

The best bases are orthonormal. Much of what follows will be for $\mathbb{F}^n$ in the interest of generality.

Definition 13.2.5 *Suppose* $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ *is a set of vectors in* $\mathbb{F}^n$. *It is an orthonormal set if*

$$\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Every orthonormal set of vectors is automatically linearly independent.

Proposition 13.2.6 *Suppose* $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ *is an orthonormal set of vectors. Then it is linearly independent.*

Proof: Suppose $\sum_{i=1}^{k} c_i \mathbf{v}_i = \mathbf{0}$. Then taking dot products with $\mathbf{v}_j$,

$$0 = \mathbf{0} \cdot \mathbf{v}_j = \sum_i c_i \mathbf{v}_i \cdot \mathbf{v}_j = \sum_i c_i \delta_{ij} = c_j.$$

Since j is arbitrary, this shows the set is linearly independent as claimed. ■

It turns out that if X is any subspace of $\mathbb{F}^m$, then there exists an orthonormal basis for X This follows from the use of the next lemma applied to a basis for X.

Lemma 13.2.7 *Let* $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ *be a linearly independent subset of* $\mathbb{F}^p$, $p \geq n$. *Then there exist orthonormal vectors* $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ *which have the property that for each* $k \leq n$, *span* $(\mathbf{x}_1, \cdots, \mathbf{x}_k) =$ *span* $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$.

Proof: Let $\mathbf{u}_1 \equiv \mathbf{x}_1 / |\mathbf{x}_1|$. Thus for $k = 1$, $\mathrm{span}(\mathbf{u}_1) = \mathrm{span}(\mathbf{x}_1)$ and $\{\mathbf{u}_1\}$ is an orthonormal set. Now suppose for some $k < n$, $\mathbf{u}_1, \cdots, \mathbf{u}_k$, have been chosen such that $(\mathbf{u}_j, \mathbf{u}_l) = \delta_{jl}$ and $\mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \mathrm{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$. Then define

$$\mathbf{u}_{k+1} \equiv \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j) \mathbf{u}_j}{\left| \mathbf{x}_{k+1} - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j) \mathbf{u}_j \right|}, \tag{13.8}$$

where the denominator is not equal to zero because the $\mathbf{x}_j$ form a basis, and so

$$\mathbf{x}_{k+1} \notin \mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \mathrm{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$$

Thus by induction,

$$\mathbf{u}_{k+1} \in \mathrm{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{x}_{k+1}) = \mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k, \mathbf{x}_{k+1}).$$

Also, $\mathbf{x}_{k+1} \in \mathrm{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1})$ which is seen easily by solving 13.8 for $\mathbf{x}_{k+1}$ and it follows

$$\mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k, \mathbf{x}_{k+1}) = \mathrm{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}).$$

If $l \leq k$,

$$(\mathbf{u}_{k+1} \cdot \mathbf{u}_l) = C \left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j)(\mathbf{u}_j \cdot \mathbf{u}_l) \right) =$$

$$C \left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j) \delta_{lj} \right) = C \left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - (\mathbf{x}_{k+1} \cdot \mathbf{u}_l) \right) = 0.$$

The vectors, $\{\mathbf{u}_j\}_{j=1}^n$, generated in this way are therefore orthonormal because each vector has unit length. ■

The process by which these vectors were generated is called the Gram Schmidt process. Note that from the construction, each $\mathbf{x}_k$ is in the span of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$. In terms of matrices, this says

$$(\mathbf{x}_1 \cdots \mathbf{x}_n) = (\mathbf{u}_1 \cdots \mathbf{u}_n)\, R$$

where R is an upper triangular matrix. This is closely related to the QR factorization discussed earlier. It is called the thin QR factorization. If the Gram Schmidt process is used to enlarge $\{\mathbf{u}_1 \cdots \mathbf{u}_n\}$ to an orthonormal basis for $\mathbb{F}^m$, $\{\mathbf{u}_1 \cdots \mathbf{u}_n, \mathbf{u}_{n+1}, \cdots, \mathbf{u}_m\}$ then if Q is the matrix which has these vectors as columns and if R is also enlarged to R' by adding in rows of zeros, if necessary, to form an $m \times n$ matrix, then the above would be of the form

$$(\mathbf{x}_1 \cdots \mathbf{x}_n) = (\mathbf{u}_1 \cdots \mathbf{u}_m)\, R'$$

and you could read off the orthonormal basis for span $(\mathbf{x}_1 \cdots \mathbf{x}_n)$ by simply taking the first n columns of $Q = (\mathbf{u}_1 \cdots \mathbf{u}_m)$. This is convenient because computer algebra systems are set up to find QR factorizations.

Example 13.2.8 *Find an orthonormal basis for span* $\left(\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}\right)$.

This is really easy to do using a computer algebra system.

$$\begin{pmatrix} 1 & 2 \\ 3 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{11}\sqrt{11} & \frac{19}{506}\sqrt{11}\sqrt{46} & \frac{3}{46}\sqrt{46} \\ \frac{3}{11}\sqrt{11} & -\frac{9}{506}\sqrt{11}\sqrt{46} & \frac{1}{46}\sqrt{46} \\ \frac{1}{11}\sqrt{11} & \frac{4}{253}\sqrt{11}\sqrt{46} & -\frac{3}{23}\sqrt{46} \end{pmatrix} \begin{pmatrix} \sqrt{11} & \frac{3}{11}\sqrt{11} \\ 0 & \frac{1}{11}\sqrt{11}\sqrt{46} \\ 0 & 0 \end{pmatrix}$$

and so the desired orthonormal basis is

$$\begin{pmatrix} \frac{1}{11}\sqrt{11} \\ \frac{3}{11}\sqrt{11} \\ \frac{1}{11}\sqrt{11} \end{pmatrix}, \begin{pmatrix} \frac{19}{506}\sqrt{11}\sqrt{46} \\ -\frac{9}{506}\sqrt{11}\sqrt{46} \\ \frac{4}{253}\sqrt{11}\sqrt{46} \end{pmatrix}$$

►

13.2.3 SCHUR'S THEOREM

Every matrix is related to an upper triangular matrix in a particularly significant way. This is Schur's theorem and it is the most important theorem in the spectral theory of matrices. The important result which makes this theorem possible is the Gram Schmidt procedure of Lemma 13.2.7

Definition 13.2.9 *An* $n \times n$ *matrix* U*, is* ***unitary*** *if* $UU^* = I = U^*U$ *where* U^* *is defined to be the transpose of the conjugate of* U*. Thus* $\overline{U_{ij}} = U^*_{ji}$*. Note that every real orthogonal matrix is unitary. For* A *any matrix* A^**, just defined as the conjugate of the transpose, is called the* ***adjoint****.*

Note that if $U = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$ where the $\mathbf{v}_k$ are orthonormal vectors in $\mathbb{C}^n$, then U is unitary. This follows because the ij^{th} entry of U^*U is $\overline{\mathbf{v}_i^T}\mathbf{v}_j = \delta_{ij}$ is since the $\mathbf{v}_i$ are assumed orthonormal.

Lemma 13.2.10 *The following holds.* $(AB)^* = B^*A^*$.

Proof: From the definition and remembering the properties of complex conjugation,

$$\begin{aligned} \left((AB)^*\right)_{ji} &= \overline{(AB)_{ij}} = \overline{\sum_k A_{ik}B_{kj}} = \sum_k \overline{A_{ik}B_{kj}} \\ &= \sum_k B^*_{jk}A^*_{ki} = (B^*A^*)_{ji} \ \blacksquare \end{aligned}$$

Theorem 13.2.11 *Let A be an $n \times n$ matrix. Then there exists a unitary matrix U such that*

$$U^* AU = T, \tag{13.9}$$

where T is an upper triangular matrix having the eigenvalues of A on the main diagonal listed according to multiplicity as roots of the characteristic equation. If A is a real matrix having all real eigenvalues, then U can be chosen to be an orthogonal real matrix.

Proof: The theorem is clearly true if A is a 1×1 matrix. Just let $U = 1$, the 1×1 matrix which has entry 1. Suppose it is true for $(n-1) \times (n-1)$ matrices, $n \geq 2$ and let A be an $n \times n$ matrix. Then let $\mathbf{v}_1$ be a unit eigenvector for A. Then there exists λ_1 such that

$$A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \ |\mathbf{v}_1| = 1.$$

Extend $\{\mathbf{v}_1\}$ to a basis and then use the Gram - Schmidt process to obtain $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, an orthonormal basis of $\mathbb{C}^n$.. Let U_0 be a matrix whose i^{th} column is $\mathbf{v}_i$ so that U_0 is unitary. Consider $U_0^* AU_0$

$$U_0^* AU_0 = \begin{pmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_n^* \end{pmatrix} \begin{pmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1^* \\ \vdots \\ \mathbf{v}_n^* \end{pmatrix} \begin{pmatrix} \lambda_1\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{pmatrix}$$

Thus $U_0^* AU_0$ is of the form

$$\begin{pmatrix} \lambda_1 & \mathbf{a} \\ \mathbf{0} & A_1 \end{pmatrix}$$

where A_1 is an $n-1 \times n-1$ matrix. Now by induction, there exists an $(n-1) \times (n-1)$ unitary matrix $\widetilde{U}_1$ such that $\widetilde{U}_1^* A_1 \widetilde{U}_1 = T_{n-1}$, an upper triangular matrix. Consider

$$U_1 \equiv \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix}.$$

Then

$$U_1^* U_1 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1^* \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I_{n-1} \end{pmatrix}$$

Also

$$\begin{aligned} U_1^* U_0^* AU_0 U_1 &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1^* \end{pmatrix} \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & T_{n-1} \end{pmatrix} \equiv T \end{aligned}$$

where T is upper triangular. Then let $U = U_0 U_1$. It is clear that this is unitary because both matrices preserve distance. Therefore, so does the product and hence U. Alternatively,

$$I = U_0 U_1 U_1^* U_0^* = (U_0 U_1)(U_0 U_1)^*$$

and so, it follows that A is similar to T and that $U_0 U_1$ is unitary. Hence A and T have the same characteristic polynomials, and since the eigenvalues of T (A) (are the diagonal entries listed with multiplicity, this proves the main conclusion of the theorem. In case A is real with all real eigenvalues, the above argument can be repeated word for word using only the real dot product to show that U can be taken to be real and orthogonal. ■

As a simple consequence of the above theorem, here is an interesting lemma.

Lemma 13.2.12 *Let A be of the form*

$$A = \begin{pmatrix} P_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_s \end{pmatrix}$$

where P_k is an $m_k \times m_k$ matrix. Then

$$\det(A) = \prod_k \det(P_k).$$

Proof: Let U_k be an $m_k \times m_k$ unitary matrix such that

$$U_k^* P_k U_k = T_k$$

where T_k is upper triangular. Then letting U denote the block diagonal matrix, having the U_i as the blocks on the diagonal,

$$U = \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s \end{pmatrix}, \ U^* = \begin{pmatrix} U_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s^* \end{pmatrix}$$

and

$$\begin{pmatrix} U_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s^* \end{pmatrix} \begin{pmatrix} P_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_s \end{pmatrix} \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s \end{pmatrix} = \begin{pmatrix} T_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_s \end{pmatrix}$$

and so

$$\det(A) = \prod_k \det(T_k) = \prod_k \det(P_k).$$

Definition 13.2.13 *An $n \times n$ matrix A is called **Hermitian** if $A = A^*$. Thus a real symmetric ($A = A^T$) matrix is Hermitian.*

Recall that from Theorem 13.2.14, the eigenvalues of a real symmetric matrix are all real.

Theorem 13.2.14 *If A is an $n \times n$ Hermitian matrix, there exists a unitary matrix U such that*

$$U^* A U = D \tag{13.10}$$

where D is a real diagonal matrix. That is, D has nonzero entries only on the main diagonal and these are real. Furthermore, the columns of U are an orthonormal basis of eigenvectors for $\mathbb{C}^n$.. If A is real and symmetric, then U can be assumed to be a real orthogonal matrix and the columns of U form an orthonormal basis for $\mathbb{R}^n$.

Proof: From Schur's theorem above, there exists U unitary (real and orthogonal if A is real) such that

$$U^* AU = T$$

where T is an upper triangular matrix. Then from Lemma 13.2.10

$$T^* = (U^* AU)^* = U^* A^* U = U^* AU = T.$$

Thus $T = T^*$ and T is upper triangular. This can only happen if T is really a diagonal matrix having real entries on the main diagonal. (If $i \neq j$, one of T_{ij} or T_{ji} equals zero. But $T_{ij} = \overline{T_{ji}}$ and so they are both zero. Also $T_{ii} = \overline{T_{ii}}$.)

Finally, let

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

where the $\mathbf{u}_i$ denote the columns of U and

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

The equation, $U^* AU = D$ implies

$$\begin{aligned} AU &= \begin{pmatrix} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{pmatrix} \\ &= UD = \begin{pmatrix} \lambda_1 \mathbf{u}_1 & \lambda_2 \mathbf{u}_2 & \cdots & \lambda_n \mathbf{u}_n \end{pmatrix} \end{aligned}$$

where the entries denote the columns of AU and UD respectively. Therefore, $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ and since the matrix is unitary, the ij^{th} entry of $U^* U$ equals δ_{ij} and so

$$\delta_{ij} = \overline{\mathbf{u}}_i^T \mathbf{u}_j = \overline{\mathbf{u}_i^T \overline{\mathbf{u}}_j} = \overline{\mathbf{u}_i \cdot \mathbf{u}_j}.$$

This proves the corollary because it shows the vectors $\{\mathbf{u}_i\}$ form an orthonormal basis. In case A is real and symmetric, simply ignore all complex conjugations in the above argument. ■

13.3 LEAST SQUARE APPROXIMATION

A very important technique is that of the least square approximation.

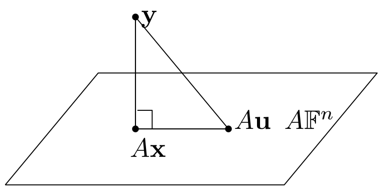
Lemma 13.3.1 *Let A be an $m \times n$ matrix and let $A(\mathbb{F}^n)$ denote the set of vectors in $\mathbb{F}^m$ which are of the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$. Then $A(\mathbb{F}^n)$ is a subspace of $\mathbb{F}^m$.*

Proof: Let $A\mathbf{x}$ and $A\mathbf{y}$ be two points of $A(\mathbb{F}^n)$. It suffices to verify that if a, b are scalars, then $aA\mathbf{x} + bA\mathbf{y}$ is also in $A(\mathbb{F}^n)$ But $aA\mathbf{x} + bA\mathbf{y} = A(a\mathbf{x} + b\mathbf{y}) \in A(\mathbb{F}^n)$ because A is linear. ■

Lemma 13.3.2 *Suppose $b \geq 0$ and $c \in \mathbb{R}$ such that $a + bt^2 + ct \geq a$ for all $t \in \mathbb{R}$, then $c = 0$.*

Proof: You need $bt^2 + ct \geq 0$ for all t. The slope of $t \mapsto bt^2 + ct$ is c when $t = 0$. Thus the inequality is violated unless $c = 0$. ■

The following theorem gives the equivalence of an orthogonality condition with a minimization condition. The following picture illustrates the geometric meaning of this theorem



Theorem 13.3.3 *Let $\mathbf{y} \in \mathbb{F}^m$ and let A be an $m \times n$ matrix. Then there exists $\mathbf{x} \in \mathbb{F}^n$ minimizing the function $\mathbf{x} \mapsto |\mathbf{y} - A\mathbf{x}|^2$. Furthermore, $\mathbf{x}$ minimizes this function if and only if*

$$((\mathbf{y} - A\mathbf{x}), A\mathbf{u}) = 0$$

for all $\mathbf{u} \in \mathbb{F}^n$.

Proof: First consider the characterization of the minimizer. Let $\mathbf{u} \in \mathbb{F}^n$. Let $|\theta| = 1$,

$$\bar{\theta}(\mathbf{y} - A\mathbf{x}, A\mathbf{u}) = |(\mathbf{y} - A\mathbf{x}, A\mathbf{u})|$$

Now consider the function of $t \in \mathbb{R}$

$$\begin{aligned} p(t) \equiv |\mathbf{y} - (A\mathbf{x} + t\theta A\mathbf{u})|^2 &= ((\mathbf{y} - A\mathbf{x}) - t\theta A\mathbf{u}, (\mathbf{y} - A\mathbf{x}) - t\theta A\mathbf{u}) \\ &= |\mathbf{y} - A\mathbf{x}|^2 + t^2 |A\mathbf{u}|^2 - 2t\mathrm{Re}(\mathbf{y} - A\mathbf{x}, \theta A\mathbf{u}) \geq |\mathbf{y} - A\mathbf{x}|^2 \\ &= |\mathbf{y} - A\mathbf{x}|^2 + t^2 |A\mathbf{u}|^2 - 2t\mathrm{Re}\bar{\theta}(\mathbf{y} - A\mathbf{x}, A\mathbf{u}) \\ &= |\mathbf{y} - A\mathbf{x}|^2 + t^2 |A\mathbf{u}|^2 - 2t|(\mathbf{y} - A\mathbf{x}, A\mathbf{u})| \end{aligned}$$

Then if $|\mathbf{y} - A\mathbf{x}|$ is as small as possible, this will occur when $t = 0$ and so $p'(0) = 0$. But this says

$$|(\mathbf{y} - A\mathbf{x}, A\mathbf{u})| = 0$$

You could also use Lemma 13.3.2 to see this is 0. Since $\mathbf{u}$ was arbitrary, this proves one direction.

Conversely, if this quantity equals 0,

$$\begin{aligned} |\mathbf{y} - (A\mathbf{x} + A\mathbf{u})|^2 &= |\mathbf{y} - A\mathbf{x}|^2 + |A\mathbf{x} - A\mathbf{u}|^2 + 2\mathrm{Re}\,(\mathbf{y} - A\mathbf{x}, A\mathbf{u}) \\ &= |\mathbf{y} - A\mathbf{x}|^2 + |A\mathbf{x} - A\mathbf{u}|^2 \end{aligned}$$

and so the minimum occurs at any point $\mathbf{z}$ such that $A\mathbf{x} = A\mathbf{z}$.

Does there exist an $\mathbf{x}$ which minimizes this function? From what was just shown, it suffices to show that there exists $\mathbf{x}$ such that $((\mathbf{y} - A\mathbf{x}), A\mathbf{u})$ for all $\mathbf{u}$. By the Gramm Schmidt process there exists an orthonormal basis $\{A\mathbf{x}_k\}$ for $A(\mathbb{F}^n)$. Then for a given $\mathbf{y}$

$$\left(\mathbf{y} - \sum_{k=1}^{r} (\mathbf{y}, A\mathbf{x}_k)\, A\mathbf{x}_k, A\mathbf{x}_j\right) = (\mathbf{y}, A\mathbf{x}_j) - \sum_{k=1}^{r} (\mathbf{y}, A\mathbf{x}_k)\, \overbrace{(A\mathbf{x}_k, A\mathbf{x}_j)}^{\delta_{kj}} = 0.$$

In particular,

$$\left(\mathbf{y}-A\left(\sum_{k=1}^{r}(\mathbf{y},A\mathbf{x}_k)\,\mathbf{x}_k\right),\mathbf{w}\right)=0$$

for all $\mathbf{w}\in A(\mathbb{F}^n)$ since $\{A\mathbf{x}_k\}$ is a basis. Therefore,

$$\mathbf{x}=\sum_{k=1}^{r}(\mathbf{y},A\mathbf{x}_k)\,\mathbf{x}_k$$

is a minimizer. So is any $\mathbf{z}$ such that $A\mathbf{z}=A\mathbf{x}$. ■

Recall the definition of the adjoint of a matrix.

Definition 13.3.4 *Let A be an $m\times n$ matrix. Then*

$$A^*\equiv\overline{(A^T)}.$$

This means you take the transpose of A and then replace each entry by its conjugate. This matrix is called the ***adjoint.*** *Thus in the case of real matrices having only real entries, the adjoint is just the transpose.*

Lemma 13.3.5 *Let A be an $m\times n$ matrix. Then*

$$A\mathbf{x}\cdot\mathbf{y}=\mathbf{x}\cdot A^*\mathbf{y}$$

Proof: This follows from the definition.

$$A\mathbf{x}\cdot\mathbf{y}=\sum_{i,j}A_{ij}x_j\overline{y_i}=\sum_{i,j}x_j\overline{A^*_{ji}}\overline{y_i}=\mathbf{x}\cdot A^*\mathbf{y}.\ ■$$

The next corollary gives the technique of least squares.

Corollary 13.3.6 *A value of $\mathbf{x}$ which solves the problem of Theorem 13.3.3 is obtained by solving the equation*

$$A^*A\mathbf{x}=A^*\mathbf{y}$$

and furthermore, there exists a solution to this system of equations.

Proof: For $\mathbf{x}$ the minimizer of Theorem 13.3.3, $(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w}=0$ for all $\mathbf{w}\in\mathbb{F}^n$ and from Lemma 13.3.5, this is the same as saying

$$A^*(\mathbf{y}-A\mathbf{x})\cdot\mathbf{w}=0$$

for all $\mathbf{w}\in\mathbb{F}^n$ This implies

$$A^*\mathbf{y}-A^*A\mathbf{x}=\mathbf{0}.$$

Therefore, there is a solution to the equation of this corollary, and it solves the minimization problem of Theorem 13.3.3. ■

Note that $\mathbf{x}$ might not be unique but $A\mathbf{x}$, the closest point of $A(\mathbb{F}^n)$. to $\mathbf{y}$ is unique. This was shown in the above argument. Sometimes people like to consider the $\mathbf{x}$ such that $A\mathbf{x}$ is as close as possible to $\mathbf{y}$ and also $|\mathbf{x}|$ is as small as possible. It turns out that there exists a unique such $\mathbf{x}$ and it is denoted as $A^+\mathbf{y}$. However, this is as far as I will go with this in this part of the book.

There is also a useful observation about orthonormal sets of vectors which is stated in the next lemma.

Lemma 13.3.7 *Suppose* $\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_r\}$ *is an orthonormal set of vectors. Then if* $c_1,\cdots,c_r$ *are scalars,*

$$\left|\sum_{k=1}^r c_k\mathbf{x}_k\right|^2=\sum_{k=1}^r|c_k|^2.$$

Proof: This follows from the definition. From the properties of the dot product and using the fact that the given set of vectors is orthonormal,

$$\left|\sum_{k=1}^r c_k\mathbf{x}_k\right|^2=\left(\sum_{k=1}^r c_k\mathbf{x}_k,\sum_{j=1}^r c_j\mathbf{x}_j\right)=\sum_{k,j}c_k\overline{c_j}\,(\mathbf{x}_k,\mathbf{x}_j)=\sum_{k=1}^r|c_k|^2.$$

13.3.1 THE LEAST SQUARES REGRESSION LINE

For the situation of the least squares regression line discussed here I will specialize to the case of $\mathbb{R}^n$ rather than $\mathbb{F}^n$ because it seems this case is by far the most interesting and the extra details are not justified by an increase in utility. Thus, everywhere you see A^* it suffices to place A^T.

An important application of Corollary 13.3.6 is the problem of finding the least squares regression line in statistics. Suppose you are given points in xy plane

$$\{(x_i, y_i)\}_{i=1}^n$$

and you would like to find constants m and b such that the line $y = mx + b$ goes through all these points. Of course this will be impossible in general. Therefore, try to find m, b to get as close as possible. The desired system is

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} \equiv A \begin{pmatrix} m \\ b \end{pmatrix}$$

which is of the form $\mathbf{y} = A\mathbf{x}$ and it is desired to choose m and b to make

$$\left| A \begin{pmatrix} m \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right|^2$$

as small as possible. According to Theorem 13.3.3 and Corollary 13.3.6, the best values for m and b occur as the solution to

$$A^T A \begin{pmatrix} m \\ b \end{pmatrix} = A^T \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}.$$

Thus, computing $A^T A$,

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

Solving this system of equations for m and b,

$$m = \frac{-\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) + \left(\sum_{i=1}^n x_i y_i\right) n}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}$$

and

$$b = \frac{-\left(\sum_{i=1}^n x_i\right) \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n y_i\right) \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}.$$

One could clearly do a least squares fit for curves of the form $y = ax^2 + bx + c$ in the same way. In this case you want to solve as well as possible for $a, b,$ and c the system

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

and one would use the same technique as above. Many other similar problems are important, including many in higher dimensions and they are all solved the same way.

13.3.2 THE FREDHOLM ALTERNATIVE

The next major result is called the Fredholm alternative. It comes from Theorem 13.3.3 and Lemma 13.3.5.

Theorem 13.3.8. *Let A be an $m \times n$ matrix. Then there exists $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$ if and only if whenever $A^*\mathbf{z} = \mathbf{0}$ it follows that $\mathbf{z} \cdot \mathbf{y} = 0$.*

Proof: First suppose that for some $\mathbf{x} \in \mathbb{F}^n$, $A\mathbf{x} = \mathbf{y}$. Then letting $A^*\mathbf{z} = \mathbf{0}$ and using Lemma 13.3.5

$$\mathbf{y} \cdot \mathbf{z} = A\mathbf{x} \cdot \mathbf{z} = \mathbf{x} \cdot A^*\mathbf{z} = \mathbf{x} \cdot \mathbf{0} = 0.$$

This proves half the theorem.

To do the other half, suppose that whenever, $A^*\mathbf{z} = \mathbf{0}$ it follows that $\mathbf{z} \cdot \mathbf{y} = 0$. It is necessary to show there exists $\mathbf{x} \in \mathbb{F}^n$ such that $\mathbf{y} = A\mathbf{x}$ From Theorem 13.3.3 there exists $\mathbf{x}$ minimizing $|\mathbf{y} - A\mathbf{x}|^2$ which therefore satisfies

$$(\mathbf{y} - A\mathbf{x}) \cdot A\mathbf{w} = 0 \tag{13.11}$$

for all $\mathbf{w} \in \mathbb{F}^n$. Therefore, for all $\mathbf{w} \in \mathbb{F}^n$,

$$A^* (\mathbf{y} - A\mathbf{x}) \cdot \mathbf{w} = 0$$

which shows that $A^* (\mathbf{y} - A\mathbf{x}) \cdot \mathbf{w} = 0$. (Why?) Therefore, by assumption,

$$(\mathbf{y} - A\mathbf{x}) \cdot \mathbf{y} = 0.$$

Now by 13.11 with $\mathbf{w} = \mathbf{x}$,

$$(\mathbf{y} - A\mathbf{x}) \cdot (\mathbf{y}-A\mathbf{x}) = (\mathbf{y} - A\mathbf{x}) \cdot \mathbf{y}- (\mathbf{y} - A\mathbf{x}) \cdot A\mathbf{x} = 0$$

showing that $\mathbf{y} = A\mathbf{x}$. ■

The following corollary is also called the Fredholm alternative.

Corollary 13.3.9 *Let A be an $m \times n$ matrix. Then A is onto if and only if A^* is one to one.*

Proof: Suppose first A is onto. Then by Theorem 13.3.8, it follows that for all $\mathbf{y} \in \mathbb{F}^m$, $\mathbf{y} \cdot \mathbf{z} = 0$ whenever $A^*\mathbf{z} = \mathbf{0}$. Therefore, let $\mathbf{y} = \mathbf{z}$ where $A^*\mathbf{z} = \mathbf{0}$ and conclude that $\mathbf{z} \cdot \mathbf{z} = 0$ whenever $A^*\mathbf{z} = \mathbf{0}$. If $A^*\mathbf{x} = A^*\mathbf{y}$, then $A^* (\mathbf{x} - \mathbf{y}) = \mathbf{0}$ and so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. Thus A^* is one to one.

Now let $\mathbf{y} \in \mathbb{F}^m$ be given. $\mathbf{y} \cdot \mathbf{z} = 0$ whenever because, since A^* is assumed to be one to one, and $\mathbf{0}$ is a solution to this equation, it must be the only solution. Therefore, by Theorem 13.3.8 there exists $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$ therefore, A is onto.

13.4 THE RIGHT POLAR FACTORIZATION

The right polar factorization involves writing a matrix as a product of two other matrices, one which preserves distances and the other which stretches and distorts. First here are some lemmas which review and add to many of the topics discussed so far about adjoints and orthonormal sets and such things. This is of fundamental significance in geometric measure theory and also in continuum mechanics. Not surprisingly the stress should depend on the part which stretches and distorts. See [8].

Lemma 13.4.1 *Let A be a Hermitian matrix such that all its eigenvalues are nonnegative. Then there exists a Hermitian matrix $A^{1/2}$ such that $A^{1/2}$ has all nonnegative eigenvalues and* $\left(A^{1/2}\right)^2 = A$.

Proof: Since A is Hermitian, there exists a diagonal matrix D having all real nonnegative entries and a unitary matrix U such that $A = U^*DU$. Then denote by $D^{1/2}$ the matrix which is obtained by replacing each diagonal entry of D with its square root. Thus $D^{1/2}D^{1/2} = D$. Then define

$$A^{1/2} \equiv U^* D^{1/2} U.$$

Then

$$\left(A^{1/2}\right)^2 = U^* D^{1/2} U U^* D^{1/2} U = U^* D U = A.$$

Since $D^{1/2}$ is real,

$$\left(U^* D^{1/2} U\right)^* = U^* \left(D^{1/2}\right)^* (U^*)^* = U^* D^{1/2} U$$

so $A^{1/2}$ is Hermitian. ■

Next it is helpful to recall the Gram Schmidt algorithm and observe a certain property stated in the next lemma.

Lemma 13.4.2 *Suppose* $\{\mathbf{w}_1, \cdots, \mathbf{w}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_p\}$ *is a linearly independent set of vectors such that* $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ *is an orthonormal set of vectors. Then when the Gram Schmidt process is applied to the vectors in the given order, it will not change any of the* $\mathbf{w}_1, \cdots, \mathbf{w}_r$.

Proof: Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ be the orthonormal set delivered by the Gram Schmidt process. Then $\mathbf{u}_1 = \mathbf{w}_1$ because by definition, $\mathbf{u}_1 \equiv \mathbf{w}_1 / |\mathbf{w}_1| = \mathbf{w}_1$. Now suppose $\mathbf{u}_j = \mathbf{w}_j$ for all $j \leq k \leq r$. Then if $k < r$, consider the definition of $\mathbf{u}_{k+1}$.

$$\mathbf{u}_{k+1} \equiv \frac{\mathbf{w}_{k+1} - \sum_{j=1}^{k+1} (\mathbf{w}_{k+1}, \mathbf{u}_j) \mathbf{u}_j}{\left|\mathbf{w}_{k+1} - \sum_{j=1}^{k+1} (\mathbf{w}_{k+1}, \mathbf{u}_j) \mathbf{u}_j\right|}$$

By induction, $\mathbf{u}_j = \mathbf{w}_j$ and so this reduces to $\mathbf{w}_{k+1} / |\mathbf{w}_{k+1}| = \mathbf{w}_{k+1}$. ■

This lemma immediately implies the following lemma.

Lemma 13.4.3 *Let* V *be a subspace of dimension* p *and let* $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ *be an orthonormal set of vectors in* V. *Then this orthonormal set of vectors may be extended to an orthonormal basis for* V,

$$\{\mathbf{w}_1, \cdots, \mathbf{w}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_p\}$$

Proof: First extend the given linearly independent set $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ to a basis for V and then apply the Gram Schmidt theorem to the resulting basis. Since $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ is orthonormal it follows from Lemma 13.4.2 the result is of the desired form, an orthonormal basis extending $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$. ■

Here is another lemma about preserving distance.

Lemma 13.4.4 *Suppose R is an $m \times n$ matrix with $m \geq n$ and R preserves distances. Then $R^* R = I$.*

Proof: Since R preserves distances, $|R\mathbf{x}| = |\mathbf{x}|$ for every $\mathbf{x}$ Therefore from the axioms of the dot product,

$$\begin{aligned} & |\mathbf{x}|^2 + |\mathbf{y}|^2 + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) = |\mathbf{x} + \mathbf{y}|^2 = (R(\mathbf{x} + \mathbf{y}), R(\mathbf{x} + \mathbf{y})) \\ = & (R\mathbf{x}, R\mathbf{x}) + (R\mathbf{y}, R\mathbf{y}) + (R\mathbf{x}, R\mathbf{y}) + (R\mathbf{y}, R\mathbf{x}) \\ = & |\mathbf{x}|^2 + |\mathbf{y}|^2 + (R^* R\mathbf{x}, \mathbf{y}) + (\mathbf{y}, R^* R\mathbf{x}) \end{aligned}$$

and so for all $\mathbf{x}, \mathbf{y}$,

$$(R^* R\mathbf{x} - \mathbf{x}, \mathbf{y}) + (\mathbf{y}, R^* R\mathbf{x} - \mathbf{x}) = 0$$

Hence for all $\mathbf{x}, \mathbf{y}$,

$$\operatorname{Re}(R^* R\mathbf{x} - \mathbf{x}, \mathbf{y}) = 0$$

Now for $\mathbf{x}, \mathbf{y}$, a given, choose $\alpha \in \mathbb{C}$ such that

$$\alpha (R^* R\mathbf{x} - \mathbf{x}, \mathbf{y}) = |(R^* R\mathbf{x} - \mathbf{x}, \mathbf{y})|$$

Then

$$0 = \operatorname{Re}(R^* R\mathbf{x} - \mathbf{x}, \overline{\alpha}\mathbf{y}) = \operatorname{Re}\alpha (R^* R\mathbf{x} - \mathbf{x}, \mathbf{y}) = |(R^* R\mathbf{x} - \mathbf{x}, \mathbf{y})|$$

Thus $|(R^* R\mathbf{x} - \mathbf{x}, \mathbf{y})| = 0$ for all $\mathbf{x}, \mathbf{y}$, because the given $\mathbf{x}, \mathbf{y}$, were arbitrary. Let $\mathbf{y} = R^* R\mathbf{x} - \mathbf{x}$ to conclude that for all $\mathbf{x}$

$$R^* R\mathbf{x} - \mathbf{x} = \mathbf{0}$$

which says $R^* R = I$ since $\mathbf{x}$ is arbitrary. ■

With this preparation, here is the big theorem about the right polar factorization.

Theorem 13.4.5 *Let F be an $m \times n$ matrix where $m \geq n$. Then there exists a Hermitian $n \times n$ matrix U which has all nonnegative eigenvalues and an $m \times n$ matrix R which preserves distances and satisfies $R^* R = I$ such that*

$$F = RU.$$

Proof: Consider F^*F. This is a Hermitian matrix because

$$(F^*F)^* = F^*(F^*)^* = F^*F$$

Also the eigenvalues of the $n \times n$ matrix F^*F are all nonnegative. This is because if $\mathbf{x}$ is an eigenvalue,

$$\lambda(\mathbf{x},\mathbf{x}) = (F^*F\mathbf{x},\mathbf{x}) = (F\mathbf{x},F\mathbf{x}) \geq 0.$$

Therefore, by Lemma 13.4.1, there exists an $n \times n$ Hermitian matrix U having all nonnegative eigenvalues such that

$$U^2 = F^*F.$$

Consider the subspace $U(\mathbb{F}^n)$. Let $\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r\}$ be an orthonormal basis for

$$U(\mathbb{F}^n) \subseteq \mathbb{F}^n.$$

Note that $U(\mathbb{F}^n)$ might not be all of $\mathbb{F}^n$ Using Lemma 13.4.3, extend to an orthonormal basis for all of $\mathbb{F}^n$,

$$\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_n\}.$$

Next observe that $\{F\mathbf{x}_1, \cdots, F\mathbf{x}_r\}$ is also an orthonormal set of vectors in $\mathbb{F}^m$. This is because

$$(F\mathbf{x}_k, F\mathbf{x}_j) = (F^*F\mathbf{x}_k, \mathbf{x}_j) = \left(U^2\mathbf{x}_k, \mathbf{x}_j\right) = (U\mathbf{x}_k, U^*\mathbf{x}_j) = (U\mathbf{x}_k, U\mathbf{x}_j) = \delta_{jk}$$

Therefore, from Lemma 13.4.3 again, this orthonormal set of vectors can be extended to an orthonormal basis for $\mathbb{F}^m$,

$$\{F\mathbf{x}_1, \cdots, F\mathbf{x}_r, \mathbf{z}_{r+1}, \cdots, \mathbf{z}_m\}$$

Thus there are at least as many $\mathbf{z}_k$ as there are $\mathbf{y}_j$. Now for $\mathbf{x} \in \mathbb{F}^n$, since

$$\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_n\}$$

is an orthonormal basis for $\mathbb{F}^n$, there exist unique scalars,

$$c_1 \cdots, c_r, d_{r+1}, \cdots, d_n$$

such that

$$\mathbf{x} = \sum_{k=1}^{r} c_k U\mathbf{x}_k + \sum_{k=r+1}^{n} d_k \mathbf{y}_k$$

Define

$$R\mathbf{x} \equiv \sum_{k=1}^{r} c_k F\mathbf{x}_k + \sum_{k=r+1}^{n} d_k \mathbf{z}_k \tag{13.12}$$

Then also there exist scalars b_k such that

$$U\mathbf{x} = \sum_{k=1}^{r} b_k U\mathbf{x}_k$$

and so from 13.12,

$$RU\mathbf{x} = \sum_{k=1}^{r} b_k F\mathbf{x}_k = F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right)$$

Is $F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) = F(\mathbf{x})$?

$$\left(F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) - F(\mathbf{x}), F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) - F(\mathbf{x})\right)$$

$$= \left((F^*F)\left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right), \left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right)\right) = \left(U^2\left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right), \left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right)\right)$$

$$= \left(U\left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right), U\left(\sum_{k=1}^{r} b_k \mathbf{x}_k - \mathbf{x}\right)\right) = \left(\sum_{k=1}^{r} b_k U\mathbf{x}_k - U\mathbf{x}, \sum_{k=1}^{r} b_k U\mathbf{x}_k - U\mathbf{x}\right) = 0$$

Therefore, $F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) = F(\mathbf{x})$ and this shows $RU\mathbf{x} = F\mathbf{x}$. From 13.12 and Lemma 13.3.7 R preserves distances. Therefore, by Lemma 13.4.4 $R^*R = I$. ■

13.5 THE SINGULAR VALUE DECOMPOSITION

In this section, A will be an $m \times n$ matrix. To begin with, here is a simple lemma.

Lemma 13.5.1 *Let A be an $m \times n$ matrix. Then A^*A is self adjoint and all its eigenvalues are nonnegative.*

Proof: It is obvious that A^*A is self adjoint. Suppose $A^*A\mathbf{x} = \lambda\mathbf{x}$. Then $\lambda|\mathbf{x}|^2 = (\lambda\mathbf{x}, \mathbf{x}) =$ Then $(A^*A\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, A\mathbf{x}) \geq 0$. ■

Definition 13.5.2 *Let A be an $m \times n$ matrix. The singular values of A are the square roots of the positive eigenvalues of A^*A*

With this definition and lemma here is the main theorem on the singular value decomposition.

Theorem 13.5.3 *Let A be an $m \times n$ matrix. Then there exist unitary matrices, U and V of the appropriate size such that*

$$U^*AV = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}$$

where σ is of the form

$$\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{pmatrix}$$

for the σ_i the singular values of A.

Proof: By the above lemma and Theorem 13.2.14 there exists an orthonormal basis, $\{\mathbf{v}_i\}_{i=1}^n$ such that $A^*A\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ where $\sigma_i^2 > 0$ for $i = 1, \cdots, k, (\sigma_i > 0)$ for and equals zero if $i > k$. Thus for $i > k$, $A\mathbf{v}_i = \mathbf{0}$ because

$$(A\mathbf{v}_i, A\mathbf{v}_i) = (A^*A\mathbf{v}_i, \mathbf{v}_i) = (\mathbf{0}, \mathbf{v}_i) = 0.$$

For $i = 1, \cdots, k$, define $\mathbf{u}_i \in \mathbb{F}^m$ by

$$\mathbf{u}_i \equiv \sigma_i^{-1}A\mathbf{v}_i.$$

Thus $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$. Now

$$\begin{aligned}(\mathbf{u}_i, \mathbf{u}_j) &= \left(\sigma_i^{-1}A\mathbf{v}_i, \sigma_j^{-1}A\mathbf{v}_j\right) = \left(\sigma_i^{-1}\mathbf{v}_i, \sigma_j^{-1}A^*A\mathbf{v}_j\right) \\ &= \left(\sigma_i^{-1}\mathbf{v}_i, \sigma_j^{-1}\sigma_j^2\mathbf{v}_j\right) = \frac{\sigma_j}{\sigma_i}(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}.\end{aligned}$$

Thus $\{\mathbf{u}_i\}_{i=1}^k$ is an orthonormal set of vectors in $\mathbb{F}^m$. Also,

$$AA^*\mathbf{u}_i = AA^*\sigma_i^{-1}A\mathbf{v}_i = \sigma_i^{-1}AA^*A\mathbf{v}_i = \sigma_i^{-1}A\sigma_i^2\mathbf{v}_i = \sigma_i^2\mathbf{u}_i.$$

Now extend $\{\mathbf{u}_i\}_{i=1}^k$ to an orthonormal basis for all of $\mathbb{F}^m, \{\mathbf{u}_i\}_{i=1}^m$ and let

$$U \equiv (\mathbf{u}_1 \cdots \mathbf{u}_m)$$

while $V \equiv (\mathbf{v}_1 \cdots \mathbf{v}_n)$. Thus U is the matrix which has the $\mathbf{u}_i$ as columns and V is defined as the matrix which has the $\mathbf{v}_i$ as columns. Then

$$U^*AV = \begin{pmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_k^* \\ \vdots \\ \mathbf{u}_m^* \end{pmatrix} A(\mathbf{v}_1 \cdots \mathbf{v}_n) = \begin{pmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_k^* \\ \vdots \\ \mathbf{u}_m^* \end{pmatrix} (\sigma_1\mathbf{u}_1 \cdots \sigma_k\mathbf{u}_k, \mathbf{0} \cdots \mathbf{0}) = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}$$

where σ is given in the statement of the theorem. ■

The singular value decomposition has as an immediate corollary the following interesting result.

Corollary 13.5.4 *Let A be an $m \times n$ matrix. Then the rank of A and A^* equals the number of singular values.*

Proof: Since V and U are unitary, it follows that

$$\begin{aligned} \operatorname{rank}(A) &= \operatorname{rank}(U^*AV) = \operatorname{rank}\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} \\ &= \text{number of singular values.} \end{aligned}$$

Also since U, V are unitary,

$$\operatorname{rank}(A^*) = \operatorname{rank}(V^*A^*U) = \operatorname{rank}\left((U^*AV)^*\right)$$
$$= \operatorname{rank}\left(\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}^*\right) = \text{number of singular values.}$$ ■

How could you go about computing the singular value decomposition? The proof of existence indicates how to do it. Here is an informal method. You have from the singular value decompositon,

$$A = U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^*, \ A^* = V \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*$$

Then it follows that

$$A^*A = V \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^* = V \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^*$$

and so $A^*AV = V \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$. Similarly, $AA^*U = U \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$. Therefore, you would find an orthonormal basis of eigenvectors for AA^* make them the columns of a matrix such that the corresponding eigenvalues are decreasing. This gives U. You could then do the same for A^*A to get V.

Example 13.5.5 *Find a singular value decomposition for the matrix*

$$A \equiv \begin{pmatrix} \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \end{pmatrix}$$

First consider A^*A

$$\begin{pmatrix} \frac{16}{5} & \frac{32}{5} & 0 \\ \frac{32}{5} & \frac{64}{5} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

What are some eigenvalues and eigenvectors? Some computing shows these are

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -\frac{2}{5}\sqrt{5} \\ \frac{1}{5}\sqrt{5} \\ 0 \end{pmatrix} \right\} \leftrightarrow 0, \left\{ \begin{pmatrix} \frac{1}{5}\sqrt{5} \\ \frac{2}{5}\sqrt{5} \\ 0 \end{pmatrix} \right\} \leftrightarrow 16$$

Thus the matrix V is given by

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

Next consider $AA^* = \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix}$. Eigenvectors and eigenvalues are

$$\left\{ \begin{pmatrix} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{pmatrix} \right\} \leftrightarrow 0, \left\{ \begin{pmatrix} \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{pmatrix} \right\} \leftrightarrow 16$$

In this case you can let U be given by

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

Lets check this. $U^*AV =$

$$\begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{2}\sqrt{5} & \frac{4}{5}\sqrt{2}\sqrt{5} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} & 0 \\ \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

This illustrates that if you have a good way to find the eigenvectors and eigenvalues for a Hermitian matrix which has nonnegative eigenvalues, then you also have a good way to find the singular value decomposition of an arbitrary matrix.

13.6 APPROXIMATION IN THE FROBENIUS NORM

The Frobenius norm is one of many norms for a matrix. It is arguably the most obvious of all norms. First here is a short discussion of the trace.

Definition 13.6.1 *Let A be an $n \times n$ matrix. Then*

$$\textit{trace}\,(A) \equiv \sum_i A_{ii}$$

just the sum of the entries on the main diagonal.

The fundamental property of the trace is in the next lemma.

Lemma 13.6.2 *Let $A = S^{-1}BS$. Then $(A) = \textit{trace}\,(B)$. Also, for any two $n \times n$ matrices A, B*

$$\textit{trace}\,(AB) = \textit{trace}\,(BA)$$

Proof: Consider the displayed formula.

$$\text{trace}\,(AB) = \sum_i \sum_j A_{ij} B_{ji}, \ \text{trace}\,(BA) = \sum_j \sum_i B_{ji} A_{ij}$$

they are the same thing. Thus if $A = S^{-1}BS.$,

$$\text{trace}\,(A) = \text{trace}\left(S^{-1}(BS)\right) = \text{trace}\left(BSS^{-1}\right) = \text{trace}\,(B). \ \blacksquare$$

Here is the definition of the Frobenius norm.

Definition 13.6.3 *Let A be a complex $m \times n$ matrix. Then*

$$||A||_F \equiv (\textit{trace}\,(AA^*))^{1/2}$$

Also this norm comes from the inner product

$$(A, B)_F \equiv \textit{trace}\,(AB^*)$$

Thus $||A||_F^2$ is easily seen to equal $\sum_{ij} |a_{ij}|^2$ so essentially, it treats the matrix as a vector in $\mathbb{F}^{m \times n}$.

Lemma 13.6.4 *Let A be an $m \times n$ complex matrix with singular matrix*

$$\Sigma = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}$$

with σ as defined above. Then

$$||\Sigma||_F^2 = ||A||_F^2 \tag{13.13}$$

and the following hold for the Frobenius norm. If U, V are unitary and of the right size,

$$||UA||_F = ||A||_F \,,\ ||UAV||_F = ||A||_F \tag{13.14}$$

Proof: From the definition and letting U, V be unitary and of the right size,

$$||UA||_F^2 \equiv \text{trace}\,(UAA^*U^*) = \text{trace}\,(AA^*) = ||A||_F^2$$

Also,

$$||AV||_F^2 \equiv \text{trace}\,(AVV^*A^*) = \text{trace}\,(AA^*) = ||A||_F^2$$

It follows

$$||UAV||_F^2 = ||AV||_F^2 = ||A||_F^2$$

Now consider 13.13. From what was just shown,

$$||A||_F^2 = ||U\Sigma V^*||_F^2 = ||\Sigma||_F^2 \ \blacksquare$$

Of course, this shows that

$$||A||_F^2 = \sum_i \sigma_i^2,$$

the sum of the squares of the singular values of A.

Why is the singular value decomposition important? It implies

$$A = U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$$

where σ is the diagonal matrix having the singular values down the diagonal. Now sometimes A is a huge matrix, 1000×2000 or something like that. This happens in applications to situations where the entries of A describe a picture. What also happens is that most of the singular values are very small. What if you deleted those which were very small, say for all $i \geq l$ and got a new matrix,

$$A' \equiv U \begin{pmatrix} \sigma' & 0 \\ 0 & 0 \end{pmatrix} V^*?$$

Then the entries of A' would end up being close to the entries of A but there is much less information to keep track of. This turns out to be very useful. More precisely, letting

$$\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{pmatrix}, \; U^*AV = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix},$$

$$||A - A'||_F^2 = \left\| U \begin{pmatrix} \sigma - \sigma' & 0 \\ 0 & 0 \end{pmatrix} V^* \right\|_F^2 = \sum_{k=l+1}^{r} \sigma_k^2$$

Thus A is approximated by A' where A' has rank $l < r$. In fact, it is also true that out of all matrices of rank l this A' is the one which is closest to A in the Frobenius norm.

Thus A is approximated by A' where A' has rank $l < r$. In fact, it is also true that out of all matrices of rank l this A' is the one which is closest to A in the Frobenius norm. Here is roughly why this is so. First consider approximating

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

as well as possible with a rank 2 matrix. It seems clear that the one which will work best is

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

More generally if σ is a $r \times r$ diagonal matrix in which the positive diagonal entries are decreasing from upper left to lower right, then the best rank l approximation to

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

would be

$$\begin{pmatrix} \sigma' & 0 \\ 0 & 0 \end{pmatrix}$$

where σ' is the upper left $l \times l$ corner of σ as in the above example.

Now suppose A is an $m \times n$ matrix. Let U, V be unitary and of the right size such that

$$U^* AV = \begin{pmatrix} \sigma_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}$$

Then suppose B approximates A as well as possible in the Frobenius norm. Then you would want

$$\|A - B\| = \|U^* AV - U^* BV\| = \left\| \begin{pmatrix} \sigma_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} - U^* BV \right\|$$

to be as small as possible. Therefore, from the above discussion, you should have

$$U^* BV = \begin{pmatrix} \sigma' & 0 \\ 0 & 0 \end{pmatrix}, B = U \begin{pmatrix} \sigma' & 0 \\ 0 & 0 \end{pmatrix} V^*$$

whereas

$$A = U \begin{pmatrix} \sigma_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} V^*$$

13.7 MOORE PENROSE INVERSE

The singular value decomposition also has a very interesting connection to the problem of least squares solutions. Recall that it was desired to find $\mathbf{x}$ such that $|A\mathbf{x} - \mathbf{y}|$ is as small as possible. Theorem 13.3.3 shows that there is a solution to this problem which can be found by solving the system $A^*A\mathbf{x} = A^*\mathbf{y}$ Each $\mathbf{x}$ which solves this system, solves the minimization problem as was shown in the lemma just mentioned. Now consider this equation for the solutions of the minimization problem in terms of the singular value decomposition.

$$\overbrace{V\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*}^{A^*}\overbrace{U\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^*}^{A}\mathbf{x} = \overbrace{V\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*}^{A^*}\mathbf{y}.$$

Therefore, this yields the following upon using block multiplication and multiplying on the left by V^*.

$$\begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^*\mathbf{x} = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y}. \tag{13.15}$$

One solution to this equation which is very easy to spot is

$$\mathbf{x} = V\begin{pmatrix} \sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y}. \tag{13.16}$$

This special $\mathbf{x}$ is denoted by $A^+\mathbf{y}$ The matrix $V\begin{pmatrix} \sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*$ is denoted by A^+. Thus $\mathbf{x}$ just defined is a solution to the least squares problem of finding the $\mathbf{x}$ such that $A\mathbf{x}$ is as close as possible to $\mathbf{y}$. Suppose now that $\mathbf{z}$ is some other solution to this least squares problem. Thus from the above,

$$\begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^*\mathbf{z} = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y}$$

and so, multiplying both sides by $\begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 0 \end{pmatrix}$,

$$\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} V^*\mathbf{z} = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y}$$

To make $V^*\mathbf{z}$ as small as possible, you would have only the first r entries of $V^*\mathbf{z}$ be nonzero since the later ones will be zeroed out anyway so they are unnecessary. Hence

$$V^*\mathbf{z} = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y}$$

and consequently

$$\mathbf{z} = V \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*\mathbf{y} \equiv A^+\mathbf{y}$$

However, minimizing $|V^*\mathbf{z}|$ is the same as minimizing $|\mathbf{z}|$ because V is unitary. Hence $A^+\mathbf{y}$ is the solution to the least squares problem which has smallest norm.

13.8 MATLAB AND SINGULAR VALUE DECOMPOSITION

MATLAB can find this very well. The syntax is [U,S,V]=svd(A) and it will give you the unitary matrices U, V such that $U^*AV = S$ where S is the singular value matrix. Here is an example.

```
A=[1,2,5;3,-2,-1];
[U,S,V]=svd(A)
```

Then press return to get the desired matrices. Check your work by typing at >> U'*A*V and press enter to see S.

MATLAB can also find the Moore Penrose inverse or pseudoinverse as follows. First enter A followed by ; and then type B=pinv(A) and press return. It will give the pseudoinverse. Here is an example where A does not have an inverse.

```
A=[1,2,3;2,4,6;-3,-2,1];
B=pinv(A)
```

13.9 EXERCISES

1. Here are some matrices. Label according to whether they are symmetric, skew symmetric, or orthogonal. If the matrix is orthogonal, determine whether it is proper or improper.

$$\text{a)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \text{b)} \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 4 \\ -3 & 4 & 7 \end{pmatrix} \quad \text{c)} \begin{pmatrix} 0 & -2 & -3 \\ 2 & 0 & -4 \\ 3 & 4 & 0 \end{pmatrix}$$

2. Show that every real matrix may be written as the sum of a skew symmetric and a symmetric matrix. **Hint:** If A is an $n \times n$ matrix, show that $B \equiv \frac{1}{2}\left(A - A^T\right)$ is skew symmetric.
3. Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$ and consider the matrix $I - \frac{2\mathbf{x}\mathbf{x}^T}{||\mathbf{x}||^2}$. Show this matrix is both symmetric and orthogonal.
4. For U an orthogonal matrix, explain why $||U\mathbf{x}|| = ||\mathbf{x}||$ for any vector $\mathbf{x}$ Next explain why if U is an $n \times n$ matrix with the property that $||U\mathbf{x}|| = ||\mathbf{x}||$ for all vectors, $\mathbf{x}$, then U must be orthogonal. Thus the orthogonal matrices are exactly those which preserve distance.

5. A quadratic form in three variables is an expression of the form $a_1x^2 + a_2y^2 + a_3z^2 + a_4xy + a_5xz + a_6yz$. Show that every such quadratic form may be written as

$$\begin{pmatrix} x & y & z \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

where A is a symmetric matrix.

6. Given a quadratic form in three variables, x, y, and z, show there exists an orthogonal matrix U and variables x', y', z' such that $\begin{pmatrix} x & y & z \end{pmatrix}^T = U \begin{pmatrix} x' & y' & z' \end{pmatrix}^T$ with the property that in terms of the new variables, the quadratic form is

$$\lambda_1 (x')^2 + \lambda_2 (y')^2 + \lambda_3 (z')^2$$

where the numbers, λ_1, λ_2, and λ_3 are the eigenvalues of the matrix A in Problem 5.

7. If A is a symmetric invertible matrix, is it always the case that A^{-1} must be symmetric also? How about A^k for k a positive integer? Explain.
8. If A, B are symmetric matrices, does it follow that AB is also symmetric?
9. Suppose A, B are symmetric and $AB = BA$. Does it follow that AB is symmetric?
10. Here are some matrices. What can you say about the eigenvalues of these matrices just by looking at them?

a) $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ c) $\begin{pmatrix} 0 & -2 & -3 \\ 2 & 0 & -4 \\ 3 & 4 & 0 \end{pmatrix}$

b) $\begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 4 \\ -3 & 4 & 7 \end{pmatrix}$ d) $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}$

11. Find the eigenvalues and eigenvectors of the matrix $\begin{pmatrix} c & 0 & 0 \\ 0 & 0 & -b \\ 0 & b & 0 \end{pmatrix}$ Here b, c are real numbers.

12. Find the eigenvalues and eigenvectors of the matrix $\begin{pmatrix} c & 0 & 0 \\ 0 & a & -b \\ 0 & b & a \end{pmatrix}$. Here a, b, c are real numbers.

13. Find the eigenvalues and an orthonormal basis of eigenvectors for A.

$$A = \begin{pmatrix} 11 & -1 & -4 \\ -1 & 11 & -4 \\ -4 & -4 & 14 \end{pmatrix}$$

Hint: Two eigenvalues are 12 and 18.

14. Find the eigenvalues and an orthonormal basis of eigenvectors for A.

$$A = \begin{pmatrix} 4 & 1 & -2 \\ 1 & 4 & -2 \\ -2 & -2 & 7 \end{pmatrix}$$

Hint: One eigenvalue is 3.

15. Show that if A is a real symmetric matrix and λ and μ are two different eigenvalues, then if $\mathbf{x}$ is an eigenvector for λ and $\mathbf{y}$ is an eigenvector for μ

then $\mathbf{x} \cdot \mathbf{y} = 0$. Also all eigenvalues are real. Supply reasons for each step in the following argument. First

$$\lambda \mathbf{x}^T \overline{\mathbf{x}} = (A\mathbf{x})^T \overline{\mathbf{x}} = \mathbf{x}^T A \overline{\mathbf{x}} = \mathbf{x}^T \overline{A\mathbf{x}} = \mathbf{x}^T \overline{\lambda} \overline{\mathbf{x}} = \overline{\lambda} \mathbf{x}^T \overline{\mathbf{x}}$$

and so $\lambda = \overline{\lambda}$. This shows that all eigenvalues are real. It follows all the eigenvectors are real. Why? Now let $\mathbf{x}, \mathbf{y}, \mu$ and λ be given as above.

$$\lambda (\mathbf{x} \cdot \mathbf{y}) = \lambda \mathbf{x} \cdot \mathbf{y} = A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mu \mathbf{y} = \mu (\mathbf{x} \cdot \mathbf{y}) = \mu (\mathbf{x} \cdot \mathbf{y})$$

and so

$$(\lambda - \mu) \mathbf{x} \cdot \mathbf{y} = 0.$$

Since $\lambda \neq \mu$, it follows $\mathbf{x} \cdot \mathbf{y} = 0$.

16. Suppose U is an orthogonal $n \times n$ matrix. Explain why $(U) = n$.
17. Show that if A is an Hermitian matrix and λ and μ are two different eigenvalues, then if $\mathbf{x}$ is an eigenvector for λ and $\mathbf{y}$ is an eigenvector for μ then $\mathbf{x} \cdot \mathbf{y} = 0$. Also all eigenvalues are real. Supply reasons for each step in the following argument. First

$$\lambda \mathbf{x} \cdot \mathbf{x} = A\mathbf{x} \cdot \mathbf{x} = \mathbf{x} \cdot A\mathbf{x} = \mathbf{x} \cdot \lambda \mathbf{x} = \overline{\lambda} \mathbf{x} \cdot \mathbf{x}$$

and so $\lambda = \overline{\lambda}$. This shows that all eigenvalues are real. Now let $\mathbf{x}, \mathbf{y}, \mu$ and λ be given as above.

$$\lambda (\mathbf{x} \cdot \mathbf{y}) = \lambda \mathbf{x} \cdot \mathbf{y} = A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mu \mathbf{y} = \overline{\mu} (\mathbf{x} \cdot \mathbf{y}) = \mu (\mathbf{x} \cdot \mathbf{y})$$

and so $(\lambda - \mu) \mathbf{x} \cdot \mathbf{y} = 0$ Since $\lambda \neq \mu$, it follows $\mathbf{x} \cdot \mathbf{y} = 0$.

18. Show that the eigenvalues and eigenvectors of a real matrix occur in conjugate pairs.
19. If a real matrix A has all real eigenvalues, does it follow that A must be symmetric. If so, explain why and if not, give an example to the contrary.
20. Suppose A is a 3×3 symmetric matrix and you have found two eigenvectors which form an orthonormal set. Explain why their cross product is also an eigenvector.
21. Study the definition of an orthonormal set of vectors. Write it from memory.
22. Determine which of the following sets of vectors are orthonormal sets. Justify your answer.

a) $\{(1,1),(1,-1)\}$

b) $\left\{\left(\frac{1}{\sqrt{2}},\frac{-1}{\sqrt{2}}\right),(1,0)\right\}$

c) $\left\{\left(\frac{1}{3},\frac{2}{3},\frac{2}{3}\right),\left(\frac{-2}{3},\frac{-1}{3},\frac{2}{3}\right),\left(\frac{2}{3},\frac{-2}{3},\frac{1}{3}\right)\right\}$

23. Show that if $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$ is an orthonormal set of vectors in $\mathbb{F}^n$, then it is a basis. **Hint:** It was shown earlier that this is a linearly independent set. If you wish, replace $\mathbb{F}^n$ with $\mathbb{R}^n$. Do this version if you do not know the dot product for vectors in $\mathbb{C}^n$.

24. Fill in the missing entries to make the matrix orthogonal.

$$\begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \_ & \_ \\ \_ & \frac{\sqrt{6}}{3} & \_ \end{pmatrix}.$$

25. Fill in the missing entries to make the matrix orthogonal.

$$\begin{pmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{1}{6}\sqrt{2} \\ \frac{2}{3} & \_ & \_ \\ \_ & 0 & \_ \end{pmatrix}$$

26. Fill in the missing entries to make the matrix orthogonal.

$$\begin{pmatrix} \frac{1}{3} & -\frac{2}{\sqrt{5}} & \_ \\ \frac{2}{3} & 0 & \_ \\ \_ & \_ & \frac{4}{15}\sqrt{5} \end{pmatrix}$$

27. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

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

Hint: One eigenvalue is -2.

28. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} 17 & -7 & -4 \\ -7 & 17 & -4 \\ -4 & -4 & 14 \end{pmatrix}.$$

Hint: Two eigenvalues are 18 and 24.

29. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} 13 & 1 & 4 \\ 1 & 13 & 4 \\ 4 & 4 & 10 \end{pmatrix}.$$

Hint: Two eigenvalues are 12 and 18.

30. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} -\frac{5}{3} & \frac{1}{15}\sqrt{6}\sqrt{5} & \frac{8}{15}\sqrt{5} \\ \frac{1}{15}\sqrt{6}\sqrt{5} & -\frac{14}{5} & -\frac{1}{15}\sqrt{6} \\ \frac{8}{15}\sqrt{5} & -\frac{1}{15}\sqrt{6} & \frac{7}{15} \end{pmatrix}$$

Hint: The eigenvalues are $-3, -2, 1$.

31. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}.$$

32. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & 1 & 5 \end{pmatrix}.$$

33. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} \frac{4}{3} & \frac{1}{3}\sqrt{3}\sqrt{2} & \frac{1}{3}\sqrt{2} \\ \frac{1}{3}\sqrt{3}\sqrt{2} & 1 & -\frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{2} & -\frac{1}{3}\sqrt{3} & \frac{5}{3} \end{pmatrix}$$

Hint: The eigenvalues are $0, 2, 2$ where 2 is listed twice because it is a root of multiplicity 2.

34. Find the eigenvalues and an orthonormal basis of eigenvectors for A. Diagonalize A by finding an orthogonal matrix U and a diagonal matrix D such that $U^T AU = D$.

$$A = \begin{pmatrix} 1 & \frac{1}{6}\sqrt{3}\sqrt{2} & \frac{1}{6}\sqrt{3}\sqrt{6} \\ \frac{1}{6}\sqrt{3}\sqrt{2} & \frac{3}{2} & \frac{1}{12}\sqrt{2}\sqrt{6} \\ \frac{1}{6}\sqrt{3}\sqrt{6} & \frac{1}{12}\sqrt{2}\sqrt{6} & \frac{1}{2} \end{pmatrix}$$

Hint: The eigenvalues are $2, 1, 0$.

35. Find the eigenvalues and an orthonormal basis of eigenvectors for the matrix

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{6}\sqrt{3}\sqrt{2} & -\frac{7}{18}\sqrt{3}\sqrt{6} \\ \frac{1}{6}\sqrt{3}\sqrt{2} & \frac{3}{2} & -\frac{1}{12}\sqrt{2}\sqrt{6} \\ -\frac{7}{18}\sqrt{3}\sqrt{6} & -\frac{1}{12}\sqrt{2}\sqrt{6} & -\frac{5}{6} \end{pmatrix}$$

Hint: The eigenvalues are $1, 2, -2$.

36. Find the eigenvalues and an orthonormal basis of eigenvectors for the matrix

$$\begin{pmatrix} -\frac{1}{2} & -\frac{1}{5}\sqrt{6}\sqrt{5} & \frac{1}{10}\sqrt{5} \\ -\frac{1}{5}\sqrt{6}\sqrt{5} & \frac{7}{5} & -\frac{1}{5}\sqrt{6} \\ \frac{1}{10}\sqrt{5} & -\frac{1}{5}\sqrt{6} & -\frac{9}{10} \end{pmatrix}$$

Hint: The eigenvalues are $-1, 2, -1$ where -1 is listed twice because it has multiplicity 2 as a zero of the characteristic equation.

37. Explain why a matrix A is symmetric if and only if there exists an orthogonal matrix U such that $A = U^T DU$ for D a diagonal matrix.
38. The proof of Theorem 13.3.3 concluded with the following observation. If $-ta + t^2 b \geq 0$ for all $t \in \mathbb{R}$ and $b \geq 0$, and then $a = 0$. Why is this so?
39. Using Schur's theorem, show that whenever A is an $n \times n$ matrix, $\det(A)$ equals the product of the eigenvalues of A.

40. In the proof of Theorem 13.3.8 the following argument was used. If $\mathbf{x} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in \mathbb{R}^n$, then $\mathbf{x} = \mathbf{0}$. Why is this so?
41. Using Corollary 13.3.9 show that a real $m \times n$ matrix is onto if and only if its transpose is one to one.
42. Suppose A is a 3×2 matrix. Is it possible that A^T is one to one? What does this say about A being onto? Prove your answer.
43. Find the least squares solution to the system $x + 2y = 1, 2x + 3y = 2, 3x + 5y = 4$.
44. You are doing experiments and have obtained the ordered pairs,

$$(0, 1), (1, 2), (2, 3.5), (3, 4)$$

Find m and b such that $y = mx + b$ approximates these four points as well as possible. Now do the same thing for $y = ax^2 + bx + c$, finding a, b, and c to give the best approximation.

45. Suppose you have several ordered triples, (x_i, y_i, z_i) Describe how to find a polynomial,

$$z = a + bx + cy + dxy + ex^2 + fy^2$$

for example giving the best fit to the given ordered triples. Is there any reason you have to use a polynomial? Would similar approaches work for other combinations of functions just as well?

46. Find an orthonormal basis for the spans of the following sets of vectors.

a) $(3,-4,0),(7,-1,0),(1,7,1)$.

b) $(3,0,-4),(11,0,2),(1,1,7)$

c) $(3,0,-4),(5,0,10),(-7,1,1)$

47. Using the Gram Schmidt process or the QR factorization, find an orthonormal basis for the span of the vectors, $(1,2,1),(2,-1,3)$ and $(1,0,0)$.
48. Using the Gram Schmidt process or the QR factorization, find an orthonormal basis for the span of the vectors, $(1,2,1,0),(2,-1,3,1)$ and $(1,0,0,1)$.
49. The set, $V \equiv \{(x,y,z) : 2x+3y-z=0\}$ is a subspace of $\mathbb{R}^3$. Find an orthonormal basis for this subspace.
50. The two level surfaces, $2x+3y-z+w=0$ and $3x-y+z+2w=0$ intersect in a subspace of $\mathbb{R}^4$, find a basis for this subspace. Next find an orthonormal basis for this subspace.
51. Let AB be a $m \times n$ matrices. Define an inner product on the set of $m \times n$ matrices by

$$(A,B)_F \equiv \text{trace}\,(AB^*) .$$

Show this is an inner product satisfying all the inner product axioms. Recall for M an $n \times n$ matrix, $(M) \equiv \sum_{i=1}^n M_{ii}$. The resulting norm, $||\cdot||_F$ is called the Frobenius norm and it can be used to measure the distance between two matrices.

52. Let A be an $m \times n$ matrix. Show $||A||_F^2 \equiv (A,A)_F = \sum_j \sigma_j^2$ where the σ_j are the singular values of A.
53. The trace of an $n \times n$ matrix M is defined as $\sum_i M_{ii}$. In other words it is the sum of the entries on the main diagonal. If A,B are $n \times n$ matrices, show $(AB) = \text{trace}\,(BA)$. Now explain why if $A = S^{-1}BS$ it follows $\text{trace}\,(A) = \text{trace}\,(B)$. **Hint:** For the first part, write these in terms of components of the matrices and it just falls out.
54. Using Problem 53 and Schur's theorem, show that the trace of an $n \times n$ matrix equals the sum of the eigenvalues.
55. If A is a general $n \times n$ matrix having possibly repeated eigenvalues, show there is a sequence $\{A_k\}$ of $n \times n$ matrices having distinct eigenvalues which has the

property that the ij^{th} entry of A_k converges to the ij^{th} entry of A for all ij. **Hint:** Use Schur's theorem.

56. Prove the Cayley Hamilton theorem as follows. First suppose A has a basis of eigenvectors $\{\mathbf{v}_k\}_{k=1}^n, A\mathbf{v}_k = \lambda_k \mathbf{v}_k$. Let $p(\lambda)$ be the characteristic polynomial. Show $p(A)\mathbf{v}_k = p(\lambda_k)\mathbf{v}_k = \mathbf{0}$ Then since $\{\mathbf{v}_k\}$ is a basis, it follows $p(A)\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$ and so $p(A) = 0$. Next in the general case, use Problem 55 to obtain a sequence $\{A_k\}$ of matrices whose entries converge to the entries of A such that A_k has n distinct eigenvalues and therefore by Theorem 12.1.15 A_k has a basis of eigenvectors. Therefore, from the first part and for $p_k(\lambda)$ the characteristic polynomial for A_k, it follows $p_k(A_k) = 0$. Now explain why and the sense in which $\lim_{k\to\infty} p_k(A_k) = p(A)$.

57. Show that the Moore Penrose inverse A^+ satisfies the following conditions.

$$AA^+A = A,\ \ A^+AA^+ = A^+,\ \ A^+A,\ AA^+ \text{ are Hermitian.}$$

Next show that if A_0 satisfies the above conditions, then it must be the Moore Penrose inverse and that if A is an $n \times n$ invertible matrix, then A^{-1} satisfies the above conditions. Thus the Moore Penrose inverse generalizes the usual notion of inverse but does not contradict it. **Hint:** Let

$$U^*AV = \Sigma \equiv \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}$$

and suppose

$$V^+A_0U = \begin{pmatrix} P & Q \\ R & S \end{pmatrix}$$

where P is the same size as σ. Now use the conditions to identify $P = \sigma, Q = 0$ etc.

58. Find the least squares solution to

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1+\varepsilon \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Next suppose ε is so small that all ε^2 terms are ignored by the computer but the terms of order ε are not ignored. Show the least squares equations in this case reduce to

$$\begin{pmatrix} 3 & 3+\varepsilon \\ 3+\varepsilon & 3+2\varepsilon \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a+b+c \\ a+b+(1+\varepsilon)c \end{pmatrix}.$$

Find the solution to this and compare the y values of the two solutions. Show that one of these is -2 times the other. This illustrates a problem with the technique for finding least squares solutions presented as the solutions to $A^*A\mathbf{x} = A^*\mathbf{y}$. One way of dealing with this problem is to use the QR factorization. This is illustrated in the next problem. It turns out that this helps alleviate some of the round off difficulties of the above.

59. Show that the equations $A^*A\mathbf{x} = A^*\mathbf{y}$ can be written as $R^*R\mathbf{x} = R^*Q^*\mathbf{y}$ where R is upper triangular and R^* is lower triangular. Explain how to solve this system efficiently. **Hint:** You first find $R\mathbf{x}$ and then you find $\mathbf{x}$ which will not be hard because R is upper triangular.

60. Show that $A^+ = (A^*A)^+ A^*$. Hint: You might use the description of A^+ in terms of the singular value decomposition.

61. Let $A = \begin{pmatrix} 1 & -3 & 0 \\ 3 & -1 & 0 \end{pmatrix}$ Then

$$\begin{pmatrix} -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}^T A^T A \begin{pmatrix} -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 16 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$AA^T = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$. A matrix U with $U^T AA^T U = \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}$ is $\begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$.

However,

$$\begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}^T \begin{pmatrix} 1 & -3 & 0 \\ 3 & -1 & 0 \end{pmatrix} \begin{pmatrix} -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}.$$

How can this be fixed so that you get $\begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$?

14 NUMERICAL METHODS FOR SOLVING LINEAR SYSTEMS

14.1 ITERATIVE METHODS FOR LINEAR SYSTEMS

Consider the problem of solving the equation

$$A\mathbf{x} = \mathbf{b} \tag{14.1}$$

where A is an $n \times n$ matrix. In many applications, the matrix A is huge and composed mainly of zeros. For such matrices, the method of Gauss elimination (row operations) is not a good way to solve the system because the row operations can destroy the zeros and storing all those zeros takes a lot of room in a computer. These systems are called sparse. To solve them it is common to use an iterative technique. The idea is to obtain a sequence of approximate solutions which get close to the true solution after a sufficient number of iterations.

Definition 14.1.1 *Let* $\{\mathbf{x}_k\}_{k=1}^{\infty}$ *be a sequence of vectors in* $\mathbb{F}^n$*. Say*

$$\mathbf{x}_k = \left(x_1^k, \cdots, x_n^k\right).$$

Then this sequence is said to converge to the vector $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{F}^n$*, written as*

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$$

if for each $j = 1, 2, \cdots, n$*,*

$$\lim_{k\to\infty} x_j^k = x_j.$$

In words, the sequence converges if the entries of the vectors in the sequence converge to the corresponding entries of $\mathbf{x}$*.*

Example 14.1.2 *Consider* $\mathbf{x}_k = \left(\sin(1/k), \frac{k^2}{1+k^2}, \ln\left(\frac{1+k^2}{k^2}\right)\right)$ *Find* $\lim_{k\to\infty} \mathbf{x}_k$*.*

From the above definition, this limit is the vector $(0, 1, 0)$ because

$$\lim_{k\to\infty} \sin(1/k) = 0, \ \lim_{k\to\infty} \frac{k^2}{1+k^2} = 1, \ \text{ and } \ \lim_{k\to\infty} \ln\left(\frac{1+k^2}{k^2}\right) = 0.$$

A more complete mathematical explanation is given in Linear Algebra. Linear Algebra

14.1.1 THE JACOBI METHOD

The first technique to be discussed here is the Jacobi method which is described in the following definition. In this technique, you have a sequence of vectors, $\{\mathbf{x}^k\}$ which converge to the solution to the linear system of equations and to get i^{th} the component of the $\mathbf{x}^{k+1}$, you use all the components of $\mathbf{x}^k$ except for the i^{th}. The precise description follows.

Definition 14.1.3 *The* ***Jacobi*** *iterative technique, also called the method of* ***simultaneous corrections,*** *is defined as follows. Let* $\mathbf{x}^1$ *be an initial vector, say the zero vector or some other vector. The method generates a succession of vectors,* $\mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \cdots$ *and hopefully this sequence of vectors will converge to the solution to 14.1. The vectors in this list are called* ***iterates*** *and they are obtained according to the following procedure. Letting* $A = (a_{ij})$*,*

$$a_{ii} x_i^{r+1} = -\sum_{j\neq i} a_{ij} x_j^r + b_i. \tag{14.2}$$

In terms of matrices, letting

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

The iterates are defined as

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix}$$

$$= \; - \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ a_{n1} & \cdots & a_{nn-1} & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{14.3}$$

If these iterates do converge, then the vector to which they converge will be a solution to the original system of equations.

The matrix on the left in 14.3 is obtained by retaining the main diagonal of A and setting every other entry equal to zero. The matrix on the right in 14.3 is obtained from A by setting every diagonal entry equal to zero and retaining all the other entries unchanged.

Example 14.1.4 *Use the Jacobi method to solve the system*

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

In terms of the matrices, the Jacobi iteration is of the form

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

14.2 USING MATLAB TO ITERATE

The syntax you can use to accomplish this iteration is as follows. This is for the purposes of illustration. In fact, you would not take an inverse of one of the matrices in practice.

```
D=[3 0 0 0;0 4 0 0;0 0 5 0;0 0 0 4];
O=[0 1 0 0;1 0 1 0;0 2 0 1;0 0 2 0];
d=1; x=[0;0;0;0]; b=[1;2;3;4]; k=1; F=inv(D);
while d>.0000001 & k<1000
y=-F*O*x+F*b; d=(y-x)'*(y-x); k=k+1;
x=y;
end
x
k
(((D+O)*x-b)'*((D+O)*x-b))^(1/2)
```

It is going to iterate till $|y-x|^2$ is smaller than 10^{-7} or 1000 iterations, whichever comes first. Of course $y = \mathbf{x}_{r+1}$ and $x = \mathbf{x}_r$. The next to last line which has k tells you how many iterations it took to get there and the bottom line tells you how close $\mathbf{x}$ is to solving the equation. This yields

$$x = \begin{pmatrix} .2069 \\ .3793 \\ .2759 \\ .8621 \end{pmatrix}, \ k = 14, \ 6.1753 \times 10^{-4}$$

14.2.1 THE GAUSS SEIDEL METHOD

The Gauss Seidel method differs from the Jacobi method in using x_j^{k+1} for all $j < i$ in going from $\mathbf{x}^k$ to $\mathbf{x}^{k+1}$. This is why it is called the method of successive corrections. The precise description of this method is in the following definition.

Definition 14.2.1 *The **Gauss Seidel** method, also called the **method of successive corrections** is given as follows. For $A = (a_{ij})$, the iterates for the problem $A\mathbf{x} = \mathbf{b}$ are obtained according to the formula*

$$\sum_{j=1}^{i} a_{ij} x_j^{r+1} = -\sum_{j=i+1}^{n} a_{ij} x_j^r + b_i. \tag{14.4}$$

In terms of matrices, letting

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

The iterates are defined as

$$\begin{aligned} &\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{nn-1} & a_{nn} \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix} \\ = & \; - \begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \end{aligned} \tag{14.5}$$

If $\mathbf{x}_r$ *converges to some* $\mathbf{x}$ *then this will be a solution to the original equation.*

In words, you set every entry in the original matrix which is strictly above the main diagonal equal to zero to obtain the matrix on the left. To get the matrix on the right, you set every entry of which is on or below the main diagonal equal to zero. Using the iteration procedure of 14.4 directly, the Gauss Seidel method makes use of the very latest information which is available at that stage of the computation.

The following example is the same as the example used to illustrate the Jacobi method.

Example 14.2.2 *Use the Gauss Seidel method to solve the system*

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

You can use MATLAB in the same way. You just use different matrices. As is the case with the Jacobi method, you would not invert the matrix but would use the description of the method given above. The following works fine for small systems of equations however.

```
L=[3 0 0 0;1 4 0 0;0 2 5 0;0 0 2 4];
U=[0 1 0 0;0 0 1 0;0 0 0 1;0 0 0 0];
d=1; x=[0;0;0;0]; b=[1;2;3;4]; k=1; F=inv(L);
while d>.0000001 & k<1000
y=-F*U*x+F*b; d=(y-x)'*(y-x); k=k+1;
x=y;
end
x
k
(((L+U)*x-b)'*((L+U)*x-b))^(1/2)
```

This yields

$$\begin{pmatrix} .207 \\ .3793 \\ .2759 \\ .8621 \end{pmatrix}, \ k = 8, \ 1.581 \times 10^{-4}$$

Thus it took only 8 iterations rather than 14. This is typical. The Gauss Seidel method is more complicated but tends to converge more quickly.

Now consider the following example.

Example 14.2.3 *Use the Gauss Seidel method to solve the system*

$$\begin{pmatrix} 1 & 4 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

The exact solution is given by doing row operations on the augmented matrix. When this is done the row echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & -\frac{5}{4} \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

and so the solution is

$$\begin{pmatrix} 6 \\ -\frac{5}{4} \\ 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 6.0 \\ -1.25 \\ 1.0 \\ .5 \end{pmatrix}$$

Using the same iteration scheme, you get the following.

$$1.0 \times 10^{45} \begin{pmatrix} -8.2309 \\ 2.2839 \\ -1.004 \\ .502 \end{pmatrix}, \; k = 1000, \; 9.114 \times 10^{44}$$

Thus the answer is totally useless, this after 1000 iterations. The error in approximating the solution is larger than 10^{44}. In other words, the method failed spectacularly to converge to anything. Row operations worked fine but this iterative procedure failed.

Why is the process which worked so well in the other examples not working here? A better question might be: Why does either process ever work at all? A complete answer to this question is given in more advanced linear algebra books. You can also see it in Linear Algebra.

Both iterative procedures for solving

$$A\mathbf{x} = \mathbf{b} \tag{14.6}$$

are of the form

$$B\mathbf{x}^{r+1} = -C\mathbf{x}^r + \mathbf{b}$$

where $A = B + C$. In the Jacobi procedure, the matrix C was obtained by setting the diagonal of A equal to zero and leaving all other entries the same while the matrix B was obtained by making every entry of A equal to zero other than the diagonal entries which are left unchanged. In the Gauss Seidel procedure, the matrix B was obtained from A by making every entry strictly above the main diagonal equal to zero and leaving the others unchanged, and C was obtained from A by making every entry on or below the main diagonal equal to zero and leaving the others unchanged. Thus in the Jacobi procedure, B is a diagonal matrix while in the Gauss Seidel procedure, B is lower triangular. Using matrices to explicitly solve for the iterates, yields

$$\mathbf{x}^{r+1} = -B^{-1}C\mathbf{x}^r + B^{-1}\mathbf{b}. \tag{14.7}$$

Theorem 14.2.4 *Let $A = B + C$ and suppose all eigenvalues of $B^{-1}C$ have absolute value less than 1 where $A = B + C$. Then the iterates in 14.7 converge to the unique solution of 14.6*

A complete explanation of this important result is found in more advanced linear algebra books. You can also see it in Linear Algebra. It depends on a theorem of Gelfand which is completely proved in this reference. Theorem 14.2.4 is very remarkable because it gives an algebraic condition for convergence, which is essentially an analytical question.

14.3 THE OPERATOR NORM*

Recall that for $\mathbf{x} \in \mathbb{C}^n$,

$$|\mathbf{x}| \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Also recall Theorem 3.2.17 which says that

$$|\mathbf{z}| \geq 0 \text{ and } |\mathbf{z}| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \tag{14.8}$$
$$\text{If } \alpha \text{ is a scalar, } |\alpha \mathbf{z}| = |\alpha| \, |\mathbf{z}| \tag{14.9}$$
$$|\mathbf{z} + \mathbf{w}| \leq |\mathbf{z}| + |\mathbf{w}| \, . \tag{14.10}$$

If you have the above axioms holding for $\|\cdot\|$ replacing $|\cdot|$ then $\|\cdot\|$ is called a norm. For example, you can easily verify that

$$\|\mathbf{x}\| \equiv \max \{|x_i| \, , i = 1, \cdots , n : \mathbf{x} = (x_1, \cdots , x_n)\}$$

is a norm. However, there are many other norms.

One important observation is that $\mathbf{x} \mapsto \|\mathbf{x}\|$ is a continuous function. This follows from the observation that from the triangle inequality,

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\| &\geq \|\mathbf{x}\| \\ \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{x}\| &= \|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x}\| \geq \|\mathbf{y}\| \end{aligned}$$

Hence

$$\begin{aligned} \|\mathbf{x}\| - \|\mathbf{y}\| &\leq \|\mathbf{x} - \mathbf{y}\| \\ \|\mathbf{y}\| - \|\mathbf{x}\| &\leq \|\mathbf{x} - \mathbf{y}\| \end{aligned}$$

and so

$$|\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\|$$

This section will involve some analysis. If you want to talk about norms, this is inevitable. It will need some of the theorems of calculus which are usually neglected. In particular, it needs the following result which is a case of the Heine Borel theorem. To see this proved, see any good calculus book, not most of the ones which are used in beginning courses on calculus.

Theorem 14.3.1 *Let S denote the points $\mathbf{x} \in \mathbb{F}^n$ such that $|\mathbf{x}| = 1$. Then if $\{\mathbf{x}_k\}_{k=1}^{\infty}$ is any sequence of points of S, there exists a subsequence which converges to a point of S.*

Definition 14.3.2 *Let A be $m \times n$ an matrix. Let $\|\cdot\|_k$ denote a norm on $\mathbb{C}^k$. Then the operator norm is defined as follows.*

$$\|A\| \equiv \max\{\|A\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\}$$

Lemma 14.3.3 *The operator norm is well defined and is in fact a norm on the vector space of $m \times n$ matrices.*

Proof: It has already been observed that the $m \times n$ matrices form a vector space starting on Page 125. Why is $\|A\| < \infty$?

claim: There exists $c > 0$ such that whenever $\|\mathbf{x}\| \leq 1$, it follows that $|\mathbf{x}| \leq c$.

Proof of the claim: If not, then there exists $\{\mathbf{x}_k\}$ such that $\|\mathbf{x}_k\| \leq 1$ but $|\mathbf{x}_k| > k$ for $k = 1, 2, \cdots$. Then $|\mathbf{x}_k / |\mathbf{x}_k|| = 1$ and so by the Heine Borel theorem from calculus, there exists a further subsequence, still denoted by k such that

$$\left|\frac{\mathbf{x}_k}{|\mathbf{x}_k|} - \mathbf{y}\right| \to 0, \ |\mathbf{y}| = 1.$$

Letting

$$\frac{\mathbf{x}_k}{|\mathbf{x}_k|} = \sum_{i=1}^{n} a_i^k \mathbf{e}_i, \ \mathbf{y} = \sum_{i=1}^{n} a_i \mathbf{e}_i,$$

It follows that $\mathbf{a}^k \to \mathbf{a}$ in $\mathbb{F}^n$. Hence

$$\left\|\frac{\mathbf{x}_k}{|\mathbf{x}_k|} - \mathbf{y}\right\| \leq \sum_{i=1}^{n} \left|a_i^k - a_i\right| \|\mathbf{e}_i\|$$

which converges to 0. However,

$$\left\|\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\right\| \leq \frac{1}{k}$$

and so, by continuity of $\|\cdot\|$ mentioned above,

$$\|\mathbf{y}\| = \lim_{k \to \infty} \left\|\frac{\mathbf{x}_k}{|\mathbf{x}_k|}\right\| = 0$$

Therefore, $\mathbf{y} = \mathbf{0}$ but also $|\mathbf{y}| = 1$, a contradiction. This proves the claim.

Now consider why $\|A\| < \infty$. Let c be as just described in the claim.

$$\sup\left\{\|A\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\right\} \leq \sup\left\{\|A\mathbf{x}\|_m : |\mathbf{x}| \leq c\right\}$$

Consider for $\mathbf{x}, \mathbf{y}$ with $|\mathbf{x}|, |\mathbf{y}| \leq c$

$$\|A\mathbf{x} - A\mathbf{y}\| = \left\|\sum_i A_{ij}\left(x_j - y_j\right)\mathbf{e}_i\right\|$$
$$\leq \sum_i |A_{ij}|\,|x_j - y_j|\,\|\mathbf{e}_i\| \leq C\,|\mathbf{x} - \mathbf{y}|$$

for some constant C. So $\mathbf{x} \mapsto A\mathbf{x}$ is continuous. Since the norm $\|\cdot\|_m$ is continuous also, it follows from the extreme value theorem of calculus that $\|A\mathbf{x}\|_m$ achieves its maximum on the compact set $\{\mathbf{x} : |\mathbf{x}| \leq c\}$. Thus $\|A\|$ is well defined. The only other issue of significance is the triangle inequality. However,

$$\begin{aligned} \|A + B\| &\equiv \max\left\{\|(A + B)\,\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\right\} \\ &\leq \max\left\{\|A\mathbf{x}\|_m + \|B\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\right\} \end{aligned}$$
$$\leq \max\left\{\|A\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\right\} + \max\left\{\|B\mathbf{x}\|_m : \|\mathbf{x}\|_n \leq 1\right\}$$
$$= \|A\| + \|B\|$$

Obviously $\|A\| = 0$ if and only if $A = 0$. The rule for scalars is also immediate. ■

The operator norm is one way to describe the magnitude of a matrix. Earlier the Frobenius norm was discussed. The Frobenius norm is actually not used as much as the operator norm. Recall that the Frobenius norm involved considering the $m \times n$ matrix as a vector in $\mathbb{F}^{mn}$ and using the usual Euclidean norm. It can be shown that it really doesn't matter which norm you use in terms of estimates because they are all equivalent. This is discussed in Problem 25 below for those who have had a legitimate calculus course, not just the usual undergraduate offering.

14.4 THE CONDITION NUMBER*

Let A be an $m \times n$ matrix and consider the problem $A\mathbf{x} = \mathbf{b}$ where it is assumed there is a unique solution to this problem. How does the solution change if A is changed a little bit and if $\mathbf{b}$ is changed a little bit? This is clearly an interesting question because you often do not know A and $\mathbf{b}$ exactly. If a small change in these quantities results in a large change in the solution $\mathbf{x}$, then it seems clear this would be undesirable. In what follows $||\cdot||$ when applied to a matrix will always refer to the operator norm.

Lemma 14.4.1 *Let A, B be $m \times n$ matrices. Then for $||\cdot||$ denoting the operator norm,*

$$||AB|| \leq ||A|| \, ||B|| \, .$$

Proof: This follows from the definition. Letting $||\mathbf{x}|| \leq 1$, it follows from the definition of the operator norm that

$$||AB\mathbf{x}|| \leq ||A|| \, ||B\mathbf{x}|| \leq ||A|| \, ||B|| \, ||\mathbf{x}|| \leq ||A|| \, ||B||$$

and so

$$||AB|| \equiv \sup_{||\mathbf{x}|| \leq 1} ||AB\mathbf{x}|| \leq ||A|| \, ||B|| \, . \blacksquare$$

Lemma 14.4.2 *Let A, B be $m \times n$ matrices such that A^{-1} exists, and suppose $||B|| < 1/\left|\left|A^{-1}\right|\right|$. Then $(A+B)^{-1}$ exists and*

$$\left|\left|(A+B)^{-1}\right|\right| \leq \left|\left|A^{-1}\right|\right| \left|\frac{1}{1 - ||A^{-1}B||}\right| .$$

The above formula makes sense because $||A^{-1}B|| < 1$.

Proof: By Lemma 14.4.1,

$$||A^{-1}B|| \leq ||A^{-1}|| \, ||B|| < ||A^{-1}|| \frac{1}{||A^{-1}||} = 1$$

Suppose $(A+B)\mathbf{x} = 0$. Then $0 = A\left(I + A^{-1}B\right)\mathbf{x}$ and so since A is one to one, $\left(I + A^{-1}B\right)\mathbf{x} = \mathbf{0}$. Therefore,

$$\begin{aligned} 0 &= \left|\left|\left(I + A^{-1}B\right)\mathbf{x}\right|\right| \geq ||\mathbf{x}|| - \left|\left|A^{-1}B\mathbf{x}\right|\right| \\ &\geq ||\mathbf{x}|| - \left|\left|A^{-1}B\right|\right| \, ||\mathbf{x}|| = \left(1 - \left|\left|A^{-1}B\right|\right|\right)||\mathbf{x}|| > 0 \end{aligned}$$

a contradiction. This also shows $\left(I + A^{-1}B\right)$ is one to one. Therefore, both $(A+B)^{-1}$ and $\left(I + A^{-1}B\right)^{-1}$ exist. Hence

$$(A+B)^{-1} = \left(A\left(I + A^{-1}B\right)\right)^{-1} = \left(I + A^{-1}B\right)^{-1}A^{-1}$$

Now if

$$\mathbf{x} = \left(I + A^{-1}B\right)^{-1}\mathbf{y}$$

for $||\mathbf{y}|| \leq 1$, then

$$\left(I + A^{-1}B\right)\mathbf{x} = \mathbf{y}$$

and so

$$||\mathbf{x}||\left(1 - \left|\left|A^{-1}B\right|\right|\right) \leq \left|\left|\mathbf{x} + A^{-1}B\mathbf{x}\right|\right| \leq ||\mathbf{y}|| = 1$$

and so

$$||\mathbf{x}|| = \left|\left|\left(I + A^{-1}B\right)^{-1}\mathbf{y}\right|\right| \leq \frac{1}{1 - ||A^{-1}B||}$$

Since $||\mathbf{y}|| \leq 1$ is arbitrary, this shows

$$\left|\left|\left(I + A^{-1}B\right)^{-1}\right|\right| \leq \frac{1}{1 - ||A^{-1}B||}$$

Therefore,

$$\begin{aligned} \left|\left|(A+B)^{-1}\right|\right| &= \left|\left|\left(I + A^{-1}B\right)^{-1}A^{-1}\right|\right| \\ &\leq \left|\left|A^{-1}\right|\right| \left|\left|\left(I + A^{-1}B\right)^{-1}\right|\right| \leq \left|\left|A^{-1}\right|\right| \frac{1}{1 - ||A^{-1}B||} \end{aligned}$$ ■

Proposition 14.4.3 *Suppose A is invertible, $\mathbf{b} \neq 0$, $A\mathbf{x} = \mathbf{b}$, and $A_1\mathbf{x}_1 = \mathbf{b}_1$ where $||A - A_1|| <. 1/\left|\left|A^{-1}\right|\right|$ Then*

$$\frac{||\mathbf{x}_1 - \mathbf{x}||}{||\mathbf{x}||} \leq \frac{1}{(1 - ||A^{-1}(A_1 - A)||)} ||A|| \left|\left|A^{-1}\right|\right| \left(\frac{||A_1 - A||}{||A||} + \frac{||\mathbf{b} - \mathbf{b}_1||}{||\mathbf{b}||}\right). \tag{14.11}$$

Proof: It follows from the assumptions that

$$A\mathbf{x} - A_1\mathbf{x} + A_1\mathbf{x} - A_1\mathbf{x}_1 = \mathbf{b} - \mathbf{b}_1.$$

Hence

$$A_1(\mathbf{x} - \mathbf{x}_1) = (A_1 - A)\mathbf{x} + \mathbf{b} - \mathbf{b}_1.$$

Now $A_1 = (A + (A_1 - A))$ and so by the above lemma, A_1^{-1} exists and so

$$\begin{aligned}(\mathbf{x} - \mathbf{x}_1) &= A_1^{-1}(A_1 - A)\mathbf{x} + A_1^{-1}(\mathbf{b} - \mathbf{b}_1) \\ &= (A + (A_1 - A))^{-1}(A_1 - A)\mathbf{x} + (A + (A_1 - A))^{-1}(\mathbf{b} - \mathbf{b}_1).\end{aligned}$$

By the estimate in Lemma 14.4.2,

$$||\mathbf{x} - \mathbf{x}_1|| \leq \frac{\left|\left|A^{-1}\right|\right|}{1 - ||A^{-1}(A_1 - A)||}(||A_1 - A|| \, ||\mathbf{x}|| + ||\mathbf{b} - \mathbf{b}_1||).$$

Dividing by $||\mathbf{x}||$,

$$\frac{||\mathbf{x} - \mathbf{x}_1||}{||\mathbf{x}||} \leq \frac{\left|\left|A^{-1}\right|\right|}{1 - ||A^{-1}(A_1 - A)||}\left(||A_1 - A|| + \frac{||\mathbf{b} - \mathbf{b}_1||}{||\mathbf{x}||}\right) \tag{14.12}$$

Now $\mathbf{b} = A\mathbf{x} = A\left(A^{-1}\mathbf{b}\right)$ and so $||\mathbf{b}|| \leq ||A|| \left|\left|A^{-1}\mathbf{b}\right|\right|$ and so

Therefore, from 14.12,

$$\begin{aligned}\frac{||\mathbf{x} - \mathbf{x}_1||}{||\mathbf{x}||} &\leq \frac{\left|\left|A^{-1}\right|\right|}{1 - ||A^{-1}(A_1 - A)||}\left(\frac{||A|| \, ||A_1 - A||}{||A||} + \frac{||A|| \, ||\mathbf{b} - \mathbf{b}_1||}{||\mathbf{b}||}\right) \\ &\leq \frac{\left|\left|A^{-1}\right|\right| ||A||}{1 - ||A^{-1}(A_1 - A)||}\left(\frac{||A_1 - A||}{||A||} + \frac{||\mathbf{b} - \mathbf{b}_1||}{||\mathbf{b}||}\right) \quad \blacksquare\end{aligned}$$

This shows that the number, $\left|\left|A^{-1}\right|\right| ||A||$, controls how sensitive the relative change in the solution of $A\mathbf{x} = \mathbf{b}$ is to small changes in A and $\mathbf{b}$. This number is called the condition number. It is bad when it is large because a small relative change in $\mathbf{b}$, for example, could yield a large relative change in $\mathbf{x}$

14.5 EXERCISES

1. Solve the system

$$\begin{pmatrix} 4 & 1 & 1 \\ 1 & 5 & 2 \\ 0 & 2 & 6 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

2. Solve the system

$$\begin{pmatrix} 4 & 1 & 1 \\ 1 & 7 & 2 \\ 0 & 2 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

3. Solve the system

$$\begin{pmatrix} 5 & 1 & 1 \\ 1 & 7 & 2 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

   using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

4. Solve the system

$$\begin{pmatrix} 7 & 1 & 0 \\ 1 & 5 & 2 \\ 0 & 2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

   using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

5. Solve the system

$$\begin{pmatrix} 5 & 0 & 1 \\ 1 & 7 & 1 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \\ 3 \end{pmatrix}$$

   using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

6. Solve the system

$$\begin{pmatrix} 5 & 0 & 1 \\ 1 & 7 & 1 \\ 0 & 2 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

   using the Gauss Seidel method and the Jacobi method. Check your answer by also solving it using row operations.

7. If you are considering a system of the form $A\mathbf{x} = \mathbf{b}$ and A^{-1} does not exist, will either the Gauss Seidel or Jacobi methods work? Explain. What does this indicate about using either of these methods for finding eigenvectors for a given eigenvalue?
8. Verify that

$$\|\mathbf{x}\|_\infty \equiv \max \{|x_i|, i = 1, \cdots, n : \mathbf{x} = (x_1, \cdots, x_n)\}$$

is a norm. Next verify that

$$\|\mathbf{x}\|_1 \equiv \sum_{i=1}^{n} |x_i| \,, \ \mathbf{x} = (x_1, \cdots, x_n)$$

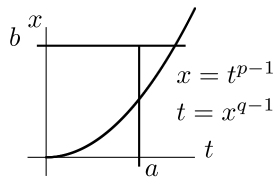
is also a norm on $\mathbb{F}^n$.

9. Let A be an $n \times n$ matrix. Denote by $\|A\|_2$ the operator norm taken with respect to the usual norm on $\mathbb{F}^n$. Show that

$$\|A\|_2 = \sigma_1$$

where σ_1 is the largest singular value. Next explain why $\left\|A^{-1}\right\|_2 = 1/\sigma_n$ where σ_n is the smallest singular value of A. Explain why the condition number reduces to σ_1/σ_n if the operator norm is defined in terms of the usual norm, $|\mathbf{x}| = \left(\sum_{j=1}^{n} |x_j|^2\right)^{1/2}$.

10. Let $p, q > 1$ and $1/p + 1/q = 1$. Consider the following picture.



Using elementary calculus, verify that for $a, b > 0$,

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab.$$

11. ↑For $p > 1$, the p norm on $\mathbb{F}^n$ is defined by

$$\|\mathbf{x}\|_p \equiv \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}$$

In fact, this is a norm and this will be shown in this and the next problem. Using the above problem in the context stated there where $p, q > 1$ and $1/p + 1/q = 1$, verify Holder's inequality

$$\sum_{k=1}^{n} |x_k| \, |y_k| \leq \|\mathbf{x}\|_p \, \|\mathbf{y}\|_q$$

Hint: You ought to consider the following.

$$\sum_{k=1}^{n} \frac{|x_k|}{\|\mathbf{x}\|_p} \frac{|y_k|}{\|\mathbf{y}\|_q}$$

Now use the result of the above problem.

12. ↑Now for $p > 1$, verify that $\|\mathbf{x}+\mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$. Then verify the other axioms of a norm. This will give an infinite collection of norms for $\mathbb{F}^n$. **Hint:** You might do the following.

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|_p^p &\leq \sum_{k=1}^n |x_k + y_k|^{p-1} (|x_k| + |y_k|) \\ &= \sum_{k=1}^n |x_k + y_k|^{p-1} |x_k| + \sum_{k=1}^n |x_k + y_k|^{p-1} |y_k|\end{aligned}$$

Now explain why $p - 1 = p/q$ and use the Holder inequality.

13. This problem will reveal the best kept secret in undergraduate mathematics, the definition of the derivative of a function of n variables. Let $\|\cdot\|$ be a norm on $\mathbb{F}^n$ and also denote by $\|\cdot\|$ a norm on $\mathbb{F}^m$. If you like, just use the standard norm on both $\mathbb{F}^n$ and $\mathbb{F}^m$. It can be shown that this doesn't matter at all (See Problem 25 on 569.) but to avoid possible confusion, you can be specific about the norm. A set $U \subseteq \mathbb{F}^n$ is said to be open if for every $\mathbf{x} \in U$, there exists some $r_{\mathbf{x}} > 0$ such that $B(\mathbf{x}, r_{\mathbf{x}}) \subseteq U$ where

$$B(\mathbf{x}, r) \equiv \{\mathbf{y} \in \mathbb{F}^n : \|\mathbf{y} - \mathbf{x}\| < r\}$$

This just says that if U contains a point $\mathbf{x}$ then it contains all the other points sufficiently near to $\mathbf{x}$. Let $\mathbf{f} : U \mapsto \mathbb{F}^m$ be a function defined on U having values in $\mathbb{F}^m$. Then $\mathbf{f}$ is differentiable at $\mathbf{x} \in U$ means that there exists an $m \times n$ matrix A such that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < \|\mathbf{v}\| < \delta$ it follows that

$$\frac{\|\mathbf{f}(\mathbf{x}+\mathbf{v}) - \mathbf{f}(\mathbf{x}) - A\mathbf{v}\|}{\|\mathbf{v}\|} < \varepsilon$$

Stated more simply,

$$\lim_{\|\mathbf{v}\| \to 0} \frac{\|\mathbf{f}(\mathbf{x}+\mathbf{v}) - \mathbf{f}(\mathbf{x}) - A\mathbf{v}\|}{\|\mathbf{v}\|} = 0$$

Show that A is unique and verify that the i^{th} column of A is

$$\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x})$$

so in particular, all partial derivatives exist. This unique $m \times n$ matrix is called the derivative of $\mathbf{f}$. It is written as $D\mathbf{f}(\mathbf{x}) = A$.

15 NUMERICAL METHODS FOR SOLVING THE EIGENVALUE PROBLEM

15.1 THE POWER METHOD FOR EIGENVALUES

This chapter presents some simple ways to find eigenvalues and eigenvectors. It is only an introduction to this important subject. However, I hope to convey some of the ideas which are used. As indicated earlier, the eigenvalue eigenvector problem is extremely difficult. Consider for example what happens if you find an eigenvalue approximately. Then you can't find an approximate eigenvector by the straight forward approach because $A - \lambda I$ is invertible whenever λ is not exactly equal to an eigenvalue.

Of course computer algebra systems allow you to ask for eigenvalues and eigenvectors and get the answer with no effort. This chapter is going to describe some of the ideas which lead to software which is able to give such answers.

The power method allows you to approximate the largest eigenvalue and also the eigenvector which goes with it. By considering the inverse of the matrix, you can also find the smallest eigenvalue. The method works in the situation of a nondefective matrix A which has a real eigenvalue of algebraic multiplicity 1, λ_n which has the property that $|\lambda_k| < |\lambda_n|$ for all $k \neq n$. Such an eigenvalue is called a dominant eigenvalue.

Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ be a basis of eigenvectors for $\mathbb{F}^n$ such that $A\mathbf{x}_n = \lambda_n \mathbf{x}_n$. Now let $\mathbf{u}_1$ be some nonzero vector. Since $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is a basis, there exists unique scalars, c_i such that

$$\mathbf{u}_1 = \sum_{k=1}^{n} c_k \mathbf{x}_k.$$

Assume you have not been so unlucky as to pick $\mathbf{u}_1$ in such a way that $c_n = 0$. Then let $A\mathbf{u}_k = \mathbf{u}_{k+1}$ so that

$$\mathbf{u}_m = A^m \mathbf{u}_1 = \sum_{k=1}^{n-1} c_k \lambda_k^m \mathbf{x}_k + \lambda_n^m c_n \mathbf{x}_n. \tag{15.1}$$

For large m the last term, $\lambda_n^m c_n \mathbf{x}_n$, determines quite well the direction of the vector on the right. This is because $|\lambda_n|$ is larger than $|\lambda_k|$ for $k < n$ and so for a large m, the sum, $\sum_{k=1}^{n-1} c_k \lambda_k^m \mathbf{x}_k$, on the right is fairly insignificant. Therefore, for large m, $\mathbf{u}_m$ is essentially a multiple of the eigenvector $\mathbf{x}_n$ the one which goes with λ_n. The only problem is that there is no control of the size of the vectors $\mathbf{u}_m$. You can fix this by scaling. Let S_2 denote the entry of $A\mathbf{u}_1$ which is largest in absolute value. We call this a **scaling factor**. Then $\mathbf{u}_2$ will not be just $A\mathbf{u}_1$ but $A\mathbf{u}_1/S_2$. Next let S_3 denote the entry of $A\mathbf{u}_2$ which has largest absolute value and define $\mathbf{u}_3 \equiv A\mathbf{u}_2/S_3$. Continue this way. The scaling just described does not destroy the relative insignificance of the term involving a sum in 15.1. Indeed it amounts to nothing more than changing the units of length. Also note that from this scaling procedure, the absolute value of the largest element of $\mathbf{u}_k$ is always equal to 1. Therefore, for large m,

$$\mathbf{u}_m = \frac{\lambda_n^m c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m} + (\text{relatively insignificant term}).$$

Therefore, the entry of $A\mathbf{u}_m$ which has the largest absolute value is essentially equal to the entry having largest absolute value of

$$A\left(\frac{\lambda_n^m c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m}\right) = \frac{\lambda_n^{m+1} c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m} \approx \lambda_n \mathbf{u}_m$$

and so for large m, it must be the case that $\lambda_n \approx S_{m+1}$. This suggests the following procedure.

Finding the largest eigenvalue with its eigenvector.

1. Start with a vector $\mathbf{u}_1$ which you hope has a component in the direction of $\mathbf{x}_n$. The vector $(1, \cdots, 1)^T$ is usually a pretty good choice.
2. If is known,

$$\mathbf{u}_{k+1} = \frac{A\mathbf{u}_k}{S_{k+1}}$$

   where S_{k+1} is the entry of $A\mathbf{u}_k$ which has largest absolute value.

3. When the scaling factors, S_k are not changing much, S_{k+1} will be close to the eigenvalue and $\mathbf{u}_{k+1}$ will be close to an eigenvector.
4. Check your answer to see if it worked well.

In finding an initial vector, it is clear that if you start with a vector which isn't too far from an eigenvector, the process will work faster. Also, the computer is able to raise the matrix to a power quite easily. You might start with $A^p\mathbf{x}$ for large p. As explained above, this will point in roughly the right direction. Then normalize it by dividing by the largest entry and use the resulting vector as your initial approximation. This ought to be close to an eigenvector and so the process would be expected to converge rapidly for this as an initial choice.

Example 15.1.1 *Find the largest eigenvalue of* $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$.

►

I will use the above suggestion.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}^{15} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.0271 \times 10^{16} \\ -5.1357 \times 10^{15} \\ 4.7018 \times 10^{11} \end{pmatrix}$$

Now divide by the largest entry to get the initial approximation for an eigenvector

$$\begin{pmatrix} 1.0271 \times 10^{16} \\ -5.1357 \times 10^{15} \\ 4.7018 \times 10^{11} \end{pmatrix} \frac{1}{1.0271 \times 10^{16}} = \begin{pmatrix} 1.0 \\ -0.50002 \\ 4.5777 \times 10^{-5} \end{pmatrix} = \mathbf{u}_1$$

The power method will now be applied to find the largest eigenvalue for the above matrix beginning with this vector.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.50002 \\ 4.5777 \times 10^{-5} \end{pmatrix} = \begin{pmatrix} 12.001 \\ -6.0003 \\ -2.5733 \times 10^{-4} \end{pmatrix}.$$

Scaling this vector by dividing by the largest entry gives

$$\begin{pmatrix} 12.001 \\ -6.0003 \\ -2.5733 \times 10^{-4} \end{pmatrix} \frac{1}{12.001} = \begin{pmatrix} 1.0 \\ -0.49998 \\ -2.1442 \times 10^{-5} \end{pmatrix} \equiv \mathbf{u}_2$$

Now lets do it again.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.49998 \\ -2.1442 \times 10^{-5} \end{pmatrix} = \begin{pmatrix} 11.999 \\ -5.9998 \\ 1.8433 \times 10^{-4} \end{pmatrix}$$

The new scaling factor is very close to the one just encountered. Therefore, it seems this is a good place to stop. The eigenvalue is approximately 11.999 and the eigenvector is close to the one obtained above. How well does it work? With the above equation, consider

$$11.999 \begin{pmatrix} 1.0 \\ -0.49998 \\ -2.1442 \times 10^{-5} \end{pmatrix} = \begin{pmatrix} 11.999 \\ -5.9993 \\ -2.5728 \times 10^{-4} \end{pmatrix}$$

These are clearly very close so this is a good approximation. In fact, the exact eigenvalue is 12 and an eigenvector is

$$\begin{pmatrix} 1.0 \\ -0.5 \\ 0 \end{pmatrix}$$

15.2 THE SHIFTED INVERSE POWER METHOD

This method can find various eigenvalues and eigenvectors. It is a significant generalization of the above simple procedure and yields very good results. The situation is this: You have a number α which is close to λ, some eigenvalue of an $n \times n$ matrix A. You don't know λ but you know that α is closer λ to than to any other eigenvalue. Your problem is to find both λ and an eigenvector which goes with λ. Another way to look at this is to start with α and seek the eigenvalue λ, which is closest to α along with an eigenvector associated with λ. If α is an eigenvalue of A, then you have what you want. Therefore, we will always assume α is not an eigenvalue of A and so $(A - \alpha I)^{-1}$ exists. When using this method it is nice to choose α fairly close to an eigenvalue. Otherwise, the method will converge slowly. In order to get some idea where to start, you could use Gerschgorin's theorem to get a rough idea where to look. The method is based on the following lemma.

Lemma 15.2.1 *Let $\{\lambda_k\}_{k=1}^n$ be the eigenvalues of A, α not an eigenvalue. Then $\mathbf{x}_k$ is an eigenvector of A for the eigenvalue λ_k, if and only if $\mathbf{x}_k$ is an eigenvector for $(A - \alpha I)^{-1}$ corresponding to the eigenvalue $\frac{1}{\lambda_k - \alpha}$.*

Proof: Let λ_k and $\mathbf{x}_k$ be as described in the statement of the lemma. Then

$$(A - \alpha I)\,\mathbf{x}_k = (\lambda_k - \alpha)\,\mathbf{x}_k$$

if and only if

$$\frac{1}{\lambda_k - \alpha}\mathbf{x}_k = (A - \alpha I)^{-1}\,\mathbf{x}_k. \quad \blacksquare$$

In explaining why the method works, we will assume A is nondefective. **This is not necessary!** One can use Gelfand's theorem on the spectral radius which is presented in [13] and invariance of $(A - \alpha I)^{-1}$ on generalized eigenspaces to prove more general results. It suffices to assume that the eigenspace for λ_k has dimension equal to the multiplicity of the eigenvalue λ_k but even this is not necessary to obtain convergence of the method. This method is better than might be supposed from the following explanation.

Pick $\mathbf{u}_1$, an initial vector and let $A\mathbf{x}_k = \lambda_k \mathbf{x}_k$ where $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is a basis of eigenvectors which exists from the assumption that A is nondefective. Assume α is closer to λ_n than to any other eigenvalue. Since A is nondefective, there exist constants, a_k such that

$$\mathbf{u}_1 = \sum_{k=1}^{n} a_k \mathbf{x}_k.$$

Possibly λ_n is a repeated eigenvalue. Then combining the terms in the sum which involve eigenvectors for λ_n, a simpler description of $\mathbf{u}_1$ is

$$\mathbf{u}_1 = \sum_{j=1}^{m} a_j \mathbf{x}_j + \mathbf{y}$$

where $\mathbf{y}$ is an eigenvector for λ_n which is assumed not equal to $\mathbf{0}$. (If you are unlucky in your choice for $\mathbf{u}_1$, this might not happen and things won't work.) Now the iteration procedure is defined as

$$\mathbf{u}_{k+1} \equiv \frac{(A - \alpha I)^{-1} \mathbf{u}_k}{S_{k+1}}$$

where S_{k+1} is the element of $(A - \alpha I)^{-1} \mathbf{u}_k$ which has largest absolute value. From Lemma 15.2.1,

$$\begin{aligned}
\mathbf{u}_{k+1} &= \frac{\sum_{j=1}^{m} a_j \left(\frac{1}{\lambda_j - \alpha}\right)^k \mathbf{x}_j + \left(\frac{1}{\lambda_n - \alpha}\right)^k \mathbf{y}}{S_2 \cdots S_{k+1}} \\
&= \frac{\left(\frac{1}{\lambda_n - \alpha}\right)^k}{S_2 \cdots S_{k+1}} \left(\sum_{j=1}^{m} a_j \left(\frac{\lambda_n - \alpha}{\lambda_j - \alpha}\right)^k \mathbf{x}_j + \mathbf{y}\right).
\end{aligned}$$

Now it is being assumed that λ_n is the eigenvalue which is closest to α and so for large k , the term,

$$\sum_{j=1}^{m} a_j \left(\frac{\lambda_n - \alpha}{\lambda_j - \alpha}\right)^k \mathbf{x}_j \equiv \mathbf{E}_k$$

is very small, while for every $k \geq 1$, $\mathbf{u}_k$ is a moderate sized vector because every entry has absolute value less than or equal to 1. Thus

$$\mathbf{u}_{k+1} = \frac{\left(\frac{1}{\lambda_n - \alpha}\right)^k}{S_2 \cdots S_{k+1}} (\mathbf{E}_k + \mathbf{y}) \equiv C_k (\mathbf{E}_k + \mathbf{y})$$

where $\mathbf{E}_k \to \mathbf{0}$, $\mathbf{y}$ is some eigenvector for λ_n, and C_k is of moderate size, remaining bounded as $k \to \infty$ due to the fact that from the construction, $\mathbf{u}_{k+1}$ has all entries no larger than 1. Therefore, for large k,

$$\mathbf{u}_{k+1} - C_k \mathbf{y} = C_k \mathbf{E}_k \approx \mathbf{0}$$

and multiplying by $(A-\alpha I)^{-1}$ yields

$$\begin{aligned}(A-\alpha I)^{-1}\mathbf{u}_{k+1}-(A-\alpha I)^{-1}C_k\mathbf{y} &= (A-\alpha I)^{-1}\mathbf{u}_{k+1}-C_k\left(\frac{1}{\lambda_n-\alpha}\right)\mathbf{y}\\ &\approx (A-\alpha I)^{-1}\mathbf{u}_{k+1}-\left(\frac{1}{\lambda_n-\alpha}\right)\mathbf{u}_{k+1}\approx \mathbf{0}.\end{aligned}$$

Therefore, for large k, $\mathbf{u}_k$ is approximately equal to an eigenvector of $(A-\alpha I)^{-1}$. Therefore,

$$(A-\alpha I)^{-1}\mathbf{u}_k \approx \frac{1}{\lambda_n-\alpha}\mathbf{u}_k$$

and so you could take the dot product of both sides with $\mathbf{u}_k$ and approximate λ_n by solving the following for λ_n.

$$\frac{(A-\alpha I)^{-1}\mathbf{u}_k\cdot\mathbf{u}_k}{|\mathbf{u}_k|^2}=\frac{1}{\lambda_n-\alpha}$$

How else can you find the eigenvalue from this? Suppose $\mathbf{u}_k=(w_1,\cdots,w_n)^T$ and from the construction $|w_i|\le 1$ and $w_k=1$ for some k. Then

$$S_{k+1}\mathbf{u}_{k+1}=(A-\alpha I)^{-1}\mathbf{u}_k\approx(A-\alpha I)^{-1}(C_{k-1}\mathbf{y})=\frac{1}{\lambda_n-\alpha}(C_{k-1}\mathbf{y})\approx\frac{1}{\lambda_n-\alpha}\mathbf{u}_k.$$

Hence the entry of $(A-\alpha I)^{-1}\mathbf{u}_k$ which has largest absolute value is approximately $\frac{1}{\lambda_n-\alpha}$ and so it is likely that you can estimate λ_n using the formula

$$S_{k+1}=\frac{1}{\lambda_n-\alpha}.$$

Of course this would fail if $(A-\alpha I)^{-1}\mathbf{u}_k$ had consistently more than one entry having equal absolute value, but this is unlikely.

Here is how you use the shifted inverse power method to find the eigenvalue and eigenvector closest to α.

1. Find $(A-\alpha I)^{-1}$.
2. Pick $\mathbf{u}_1$. It is important that $\mathbf{u}_1=\sum_{j=1}^{m}a_j\mathbf{x}_j+\mathbf{y}$ where $\mathbf{y}$ is an eigenvector which goes with the eigenvalue closest to and the sum is in an "invariant subspace corresponding to the other eigenvalues". Of course you have no way of knowing whether this is so but it typically is so. If things don't work out, just start with a different $\mathbf{u}_1$. You were phenomenally unlucky in your choice.
3. If $\mathbf{u}_k$ has been obtained,

   $$\mathbf{u}_{k+1}=\frac{(A-\alpha I)^{-1}\mathbf{u}_k}{S_{k+1}}$$

   where S_{k+1} is the element of $(A-\alpha I)^{-1}\mathbf{u}_k$ which has largest absolute value.

4. When the scaling factors, S_{k+1} are not changing much and the $\mathbf{u}_k$ are not changing much, find the approximation to the eigenvalue by solving

   $$S_{k+1}=\frac{1}{\lambda-\alpha}$$

   for λ. The eigenvector is approximated by $\mathbf{u}_{k+1}$.

5. Check your work by multiplying by the original matrix to see how well what you have found works.

Also note that this is just the power method applied to $(A-\lambda I)^{-1}$. The eigenvalue you want is the one which makes $\frac{1}{\lambda-\alpha}$ as large as possible for all $\lambda\in\sigma(A)$. This is because making $\lambda-\alpha$ small is the same as making $(\lambda-\alpha)^{-1}$ large.

15.3 AUTOMATION WITH MATLAB

You can do the above example and other examples using MATLAB. Here are some commands which will do this. It is done here for a 3×3 matrix but you adapt for any size.

```
a=[5 -8 6;1 0 0;0 1 0]; b=i; F=inv(a-b*eye(3));
S=1; u=[1;1;1]; d=1; k=1;
while d>.00001 & k<1000
w=F*u; [M,I]=max(abs(w)); T=w(I); u=w/T;
d=abs(T-S); S=T; k=k+1;
end
u
b+1/T
k
a*u-(b+1/T)*u
```

Note how the "while loop" is limited to 1000 iterations. That way it won't go on forever if there is something wrong. This asks for the eigenvalue closest to $b = i$. When MATLAB stalls, to get it to quit, you type control c. The last line checks the answer and the line with k tells the number of iterations used. Also, the funny notation [M,I]=max(abs(w)); T=w(I); gets it to pick out the entry which has largest absolute value w(I) and keep that entry unchanged. The above iteration finds the eigenvalue closest to i along with the corresponding eigenvector. When the procedure does not work well for b real, you might imagine that there are complex eigenvalues and so, since the above procedure is going to give you real approximations, it can't find the complex eigenvalues. Thus you should take b to be complex as done above.

If you have MATLAB work the above iteration, you get the following for the eigenvector eigenvalue and number of iterations, and error .

$$\begin{pmatrix} 1 \\ .5 - .5i \\ -.5i \end{pmatrix}, \ 1 + i, \ k = 18, \ 10^{-5} \begin{pmatrix} 0 \\ -0.1321 + 0.1862i \\ -0.1325 + 0.1863i \end{pmatrix}$$

In fact, this eigenvector is exactly right as is the eigenvalue $1 + i$.

Thus this method will find eigenvalues real or complex along with an eigenvector associated with the eigenvalue. Note that the characteristic polynomial of the above matrix is $\lambda^3 - 5\lambda^2 + 8\lambda - 6$ and the above finds a complex root to this polynomial. More generally, if you have a polynomial $\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$, a matrix which has this as its characteristic polynomial is called a companion matrix and you can show a matrix which works for this polynomial is of the form

$$\begin{pmatrix} -a_{n-1} & -a_{n-2} & \cdots & a_0 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}$$

Thus this method is capable of finding roots to a polynomial equation which are close to a given complex number. Of course there is a problem with determining which number you should pick. A way to determine this will be discussed later. It involves something called the QR algorithm.

Example 15.3.1 *Find the eigenvalue of* $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$ *which is closest to* -7. *Also find an eigenvector which goes with this eigenvalue.*

We use the algorithm described above.

```
a=[5 -14 11;-4 4 -4;3 6 -3]; b=-7; F=inv(a-b*eye(3));
S=1; u=[1;1;1]; d=1; k=1;
while d>.0001 & k<1000
w=F*u; [M,I]=max(abs(w)); T=w(I); u=w/T;
d=abs(T-S); S=T; k=k+1;
end
u
b+1/T
a*u-(b+1/T)*u
```

This yields

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \ -6$$

for the eigenvector and eigenvalue. In fact, this is exactly correct.

Example 15.3.2 *Consider the symmetric matrix* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}$. *Find the middle eigenvalue and an eigenvector which goes with it.*

Since A is symmetric, it follows it has three **real** eigenvalues which are solutions to

$$\begin{aligned} p(\lambda) &= \det\left(\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}\right) \\ &= \lambda^3 - 4\lambda^2 - 24\lambda - 17 = 0 \end{aligned}$$

If you use your graphing calculator to graph this polynomial, you find there is an eigenvalue somewhere between $-.9$ and $-.8$ and that this is the middle eigenvalue. Of course you could zoom in and find it very accurately without much trouble but what about the eigenvector which goes with it? If you try to solve

$$\left((-.8)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

there will be only the zero solution because the matrix on the left will be invertible and the same will be true if you replace $-.8$ with a better approximation like $-.86$ or $-.855$. This is because all these are only approximations to the eigenvalue and so the matrix in the above is nonsingular for all of these. Therefore, you will only get the zero solution and

Eigenvectors are never equal to zero!

However, there exists such an eigenvector and you can find it using the shifted inverse power method. Pick $\alpha = -.855$ in the above algorithm. Then entering the matrix and running the algorithm yields the eigenvector and eigenvalue

$$\begin{pmatrix} -.0111 \\ -.2776 \\ .2470 \end{pmatrix}, \ -.8569$$

In fact the error is on the order of 10^{-14}.

There is an easy to use trick which will eliminate some of the fuss and bother in using the shifted inverse power method. If you have

$$(A - \alpha I)^{-1} \mathbf{x} = \mu \mathbf{x}$$

then multiplying through by $(A - \alpha I)$, one finds that $\mathbf{x}$ will be an eigenvector for A with eigenvalue $\alpha + \mu^{-1}$. Hence you could simply take $(A - \alpha I)^{-1}$ to a high power and multiply by a vector to get a vector which points in the direction of an eigenvalue of A. Then divide by the largest entry and identify the eigenvalue directly by multiplying the eigenvector by A. This is illustrated in the next example.

Example 15.3.3 *Find the eigenvalues and eigenvectors of the matrix*

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix}.$$

This is only a 3×3 matrix and so it is not hard to estimate the eigenvalues. Just get the characteristic equation, graph it using a calculator and zoom in to find the eigenvalues. If you do this, you find there is an eigenvalue near -1.2, one near $-.4$, and one near 5.5. (The characteristic equation is $2 + 8\lambda + 4\lambda^2 - \lambda^3 = 0$.) Of course we have no idea what the eigenvectors are.

Lets first try to find the eigenvector and a better approximation for the eigenvalue near -1.2 . In this case, let $\alpha = -1.2$. Then

$$(A - \alpha I)^{-1} = \begin{pmatrix} -25.357143 & -33.928571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214286 & 30.357143 & -45.0 \end{pmatrix}.$$

Then

$$\begin{pmatrix} -25.357143 & -33.928571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214286 & 30.357143 & -45.0 \end{pmatrix}^{17} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} -4.9432 \times 10^{28} \\ 2.4312 \times 10^{28} \\ 4.4928 \times 10^{28} \end{pmatrix}$$

The initial approximation for an eigenvector will then be the above divided by its largest entry.

$$\begin{pmatrix} -4.9432 \times 10^{28} \\ 2.4312 \times 10^{28} \\ 4.4928 \times 10^{28} \end{pmatrix} \frac{1}{-4.9432 \times 10^{28}} = \begin{pmatrix} 1.0 \\ -0.49183 \\ -0.90888 \end{pmatrix}$$

How close is this to being an eigenvector?

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ -0.49183 \\ -0.90888 \end{pmatrix} = \begin{pmatrix} -1.2185 \\ 0.59929 \\ 1.1075 \end{pmatrix}$$
$$-1.2185 \begin{pmatrix} 1.0 \\ -0.49183 \\ -0.90888 \end{pmatrix} = \begin{pmatrix} -1.2185 \\ 0.59929 \\ 1.1075 \end{pmatrix}$$

For all practical purposes, this has found the eigenvector for the eigenvalue -1.2185.

Next we shall find the eigenvector and a more precise value for the eigenvalue near $-.4$. In this case,

$$(A - \alpha I)^{-1} = \begin{pmatrix} 8.0645161 \times 10^{-2} & -9.2741935 & 6.4516129 \\ -.40322581 & 11.370968 & -7.2580645 \\ .40322581 & 3.6290323 & -2.7419355 \end{pmatrix}.$$

The first approximation to an eigenvector can be obtained as before.

$$\begin{pmatrix} 8.0645161\times10^{-2} & -9.2741935 & 6.4516129 \\ -.40322581 & 11.370968 & -7.2580645 \\ .40322581 & 3.6290323 & -2.7419355 \end{pmatrix}^{17}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} -1.8535\times10^{16} \\ 2.3724\times10^{16} \\ 6.2874\times10^{15} \end{pmatrix}$$

The first choice for an approximate eigenvector is

$$\begin{pmatrix} -1.8535\times10^{16} \\ 2.3724\times10^{16} \\ 6.2874\times10^{15} \end{pmatrix}\frac{1}{2.3724\times10^{16}} = \begin{pmatrix} -0.78128 \\ 1.0 \\ 0.26502 \end{pmatrix}$$

Lets see how well this works as an eigenvector.

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix}\begin{pmatrix} -0.78128 \\ 1.0 \\ 0.26502 \end{pmatrix} = \begin{pmatrix} 0.2325 \\ -0.29754 \\ -0.07882 \end{pmatrix}$$
$$(-0.29754)\begin{pmatrix} -0.78128 \\ 1.0 \\ 0.26502 \end{pmatrix} = \begin{pmatrix} 0.23246 \\ -0.29754 \\ -7.8854\times10^{-2} \end{pmatrix}$$

Thus this works as an eigenvector with the eigenvalue (-0.29754).

Next we will find the eigenvalue and eigenvector for the eigenvalue near 5.5. In this case,

$$(A-\alpha I)^{-1} = \begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix}.$$

As before, I have no idea what the eigenvector is but to avoid giving the impression that you always need to start with the vector $(1,1,1)^T$, let $\mathbf{u}_1 = (1,2,3)^T$. I will use the same shortcut to get this eigenvector as in the above case.

$$\begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix}^{16}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.0987\times10^{29} \\ 7.1868\times10^{28} \\ 1.0482\times10^{29} \end{pmatrix}$$

Then dividing by the largest entry, a good guess for the eigenvector is

$$\begin{pmatrix} 1.0 \\ 0.65412 \\ 0.95404 \end{pmatrix}$$

To see if more iteration would be needed, check this.

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ 0.65412 \\ 0.95404 \end{pmatrix} = \begin{pmatrix} 5.5162 \\ 3.6082 \\ 5.2623 \end{pmatrix}$$

and

$$5.5162 \begin{pmatrix} 1.0 \\ 0.65412 \\ 0.95404 \end{pmatrix} = \begin{pmatrix} 5.5162 \\ 3.6083 \\ 5.2627 \end{pmatrix}$$

Thus this is essentially an eigenvector with eigenvalue equal to 5.5162.

15.4 THE RAYLEIGH QUOTIENT

There are many specialized results concerning the eigenvalues and eigenvectors for Hermitian matrices. A matrix A is Hermitian if $A = A^*$ where A^* means to take the transpose of the conjugate of A. In the case of a real matrix, Hermitian reduces to symmetric. Recall also that for $\mathbf{x} \in \mathbb{F}^n$,

$$|\mathbf{x}|^2 = \mathbf{x}^*\mathbf{x} = \sum_{j=1}^{n} |x_j|^2 .$$

The following corollary gives the theoretical foundation for the spectral theory of Hermitian matrices. This is a corollary of a theorem which is proved Corollary 13.2.14 and Theorem 13.2.14 on Page 428.

Corollary 15.4.1 *If A is Hermitian, then all the eigenvalues of A are real and there exists an orthonormal basis of eigenvectors.*

Thus for $\{\mathbf{x}_k\}_{k=1}^n$ this orthonormal basis,

$$\mathbf{x}_i^*\mathbf{x}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

For $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{x} \neq \mathbf{0}$, the **Rayleigh quotient** is defined by

$$\frac{\mathbf{x}^* A\mathbf{x}}{|\mathbf{x}|^2}.$$

Now let the eigenvalues of A be A be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $A\mathbf{x}_k = \lambda_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the above orthonormal basis of eigenvectors mentioned in the corollary. Then if $\mathbf{x}$ is an arbitrary vector, there exist constants, a_i such that

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{x}_i.$$

Also,

$$|\mathbf{x}|^2 = \sum_{i=1}^{n} \overline{a}_i \mathbf{x}_i^* \sum_{j=1}^{n} a_j \mathbf{x}_j = \sum_{ij} \overline{a}_i a_j \mathbf{x}_i^* \mathbf{x}_j = \sum_{ij} \overline{a}_i a_j \delta_{ij} = \sum_{i=1}^{n} |a_i|^2 .$$

Therefore,

$$\frac{\mathbf{x}^* A\mathbf{x}}{|\mathbf{x}|^2} = \frac{\left(\sum_{i=1}^n \overline{a}_i \mathbf{x}_i^*\right)\left(\sum_{j=1}^n a_j \lambda_j \mathbf{x}_j\right)}{\sum_{i=1}^n |a_i|^2}$$

$$= \frac{\sum_{ij} \overline{a}_i a_j \lambda_j \mathbf{x}_i^* \mathbf{x}_j}{\sum_{i=1}^n |a_i|^2} = \frac{\sum_{ij} \overline{a}_i a_j \lambda_j \delta_{ij}}{\sum_{i=1}^n |a_i|^2} = \frac{\sum_{i=1}^n |a_i|^2 \lambda_i}{\sum_{i=1}^n |a_i|^2} \in [\lambda_1, \lambda_n] .$$

In other words, the Rayleigh quotient is always between the largest and the smallest eigenvalues of A. When $\mathbf{x} = \mathbf{x}_n$, the Rayleigh quotient equals the largest eigenvalue and when $\mathbf{x} = \mathbf{x}_1$ the Rayleigh quotient equals the smallest eigenvalue. Suppose you calculate a Rayleigh quotient. How close is it to some eigenvalue?

Theorem 15.4.2 *Let* $\mathbf{x} \neq \mathbf{0}$ *and form the* ***Rayleigh quotient,***

$$\frac{\mathbf{x}^* A\mathbf{x}}{|\mathbf{x}|^2} \equiv q.$$

Then there exists an eigenvalue of A*, denoted here by* λ_q *such that*

$$|\lambda_q - q| \leq \frac{|A\mathbf{x} - q\mathbf{x}|}{|\mathbf{x}|}. \tag{15.2}$$

Proof: Let $\mathbf{x} = \sum_{k=1}^n a_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the orthonormal basis of eigenvectors.

$$\begin{aligned}
|A\mathbf{x} - q\mathbf{x}|^2 &= (A\mathbf{x} - q\mathbf{x})^* (A\mathbf{x} - q\mathbf{x}) \\
&= \left(\sum_{k=1}^n a_k\lambda_k\mathbf{x}_k - qa_k\mathbf{x}_k\right)^* \left(\sum_{k=1}^n a_k\lambda_k\mathbf{x}_k - qa_k\mathbf{x}_k\right) \\
&= \left(\sum_{j=1}^n (\lambda_j - q)\,\overline{a}_j\mathbf{x}_j^*\right)\left(\sum_{k=1}^n (\lambda_k - q)\,a_k\mathbf{x}_k\right) \\
&= \sum_{j,k} (\lambda_j - q)\,\overline{a}_j\,(\lambda_k - q)\,a_k\mathbf{x}_j^*\mathbf{x}_k \\
&= \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2
\end{aligned}$$

Now pick the eigenvalue, λ_q which is closest to q. Then

$$|A\mathbf{x} - q\mathbf{x}|^2 = \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2 \geq (\lambda_q - q)^2 \sum_{k=1}^n |a_k|^2 = (\lambda_q - q)^2 |\mathbf{x}|^2$$

which implies 15.2. ■

Example 15.4.3 *Consider the symmetric matrix* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}$ *Let* $\mathbf{x} = (1,1,1)^T$

How close is the Rayleigh quotient to some eigenvalue of A? *Find the eigenvector and eigenvalue to several decimal places.*

Everything is real and so there is no need to worry about taking conjugates. Therefore, the Rayleigh quotient is

$$\frac{\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}{3} = \frac{19}{3}$$

According to the above theorem, there is some eigenvalue of this matrix, λ_q such that

$$\begin{aligned}
\left|\lambda_q - \frac{19}{3}\right| &\leq \frac{\left|\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{19}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right|}{\sqrt{3}} \\
&= \frac{1}{\sqrt{3}}\begin{pmatrix} -\frac{1}{3} \\ -\frac{4}{3} \\ \frac{5}{3} \end{pmatrix} \\
&= \frac{\sqrt{\frac{1}{9} + \left(\frac{4}{3}\right)^2 + \left(\frac{5}{3}\right)^2}}{\sqrt{3}} = 1.\,247\,2
\end{aligned}$$

Could you find this eigenvalue and associated eigenvector? Of course you could. This is what the inverse shifted power method is all about.

Solve

$$\left(\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} - \frac{19}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

In other words solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and divide by the entry which is largest, 3.8707, to get

$$\mathbf{u}_2 = \begin{pmatrix} .69925 \\ .49389 \\ 1.0 \end{pmatrix}$$

Now solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .69925 \\ .49389 \\ 1.0 \end{pmatrix}$$

and divide by the entry with largest absolute value, 2.9979 to get

$$\mathbf{u}_3 = \begin{pmatrix} .71473 \\ .52263 \\ 1.0 \end{pmatrix}$$

Now solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .71473 \\ .52263 \\ 1.0 \end{pmatrix}$$

and divide by the entry with largest absolute value, 3.0454, to get

$$\mathbf{u}_4 = \begin{pmatrix} .7137 \\ .52056 \\ 1.0 \end{pmatrix}$$

Solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .7137 \\ .52056 \\ 1.0 \end{pmatrix}$$

and divide by the largest entry, 3.0421 to get

$$\mathbf{u}_5 = \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix}$$

You can see these scaling factors are not changing much. The predicted eigenvalue is obtained by solving

$$\frac{1}{\lambda - \frac{19}{3}} = 3.0421$$

to obtain $\lambda = 6.6621$. How close is this?

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.7552 \\ 3.469 \\ 6.6621 \end{pmatrix}$$

while

$$6.6621 \begin{pmatrix} .71378 \\ .52073 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.7553 \\ 3.4692 \\ 6.6621 \end{pmatrix}.$$

You see that for practical purposes, this has found the eigenvalue and an eigenvector.

15.5 THE QR ALGORITHM

15.5.1 BASIC CONSIDERATIONS

The QR algorithm is one of the most remarkable techniques for finding eigenvalues. In this section, I will discuss this method. To see more on this algorithm, consult Golub and Van Loan [6]. For an explanation of why the algorithm works see Wilkinson [18]. There is also more discussion in Linear Algebra. This will only discuss real matrices for the sake of simplicity. Also, there is a lot more to this algorithm than will be presented here. First here is an introductory lemma.

Lemma 15.5.1 *Suppose A is a block upper triangular matrix,*

$$A = \begin{pmatrix} B_1 & & * \\ & \ddots & \\ 0 & & B_r \end{pmatrix}$$

This means that the B_i are $r_i \times r_i$ matrices whose diagonals are subsets of the main diagonal of A. Then $\sigma(A) = \cup_{i=1}^{r} \sigma(B_i)$.

Proof: Say $Q_i^* B_i Q_i = T_i$ where T_i is upper triangular. Such unitary matrices exist by Schur's theorem. Then consider the similarity transformation,

$$\begin{pmatrix} Q_1^* & & 0 \\ & \ddots & \\ 0 & & Q_r^* \end{pmatrix} \begin{pmatrix} B_1 & & * \\ & \ddots & \\ 0 & & B_r \end{pmatrix} \begin{pmatrix} Q_1 & & 0 \\ & \ddots & \\ 0 & & Q_r \end{pmatrix}$$

By block multiplication this equals

$$\begin{pmatrix} Q_1^* & & 0 \\ & \ddots & \\ 0 & & Q_r^* \end{pmatrix} \begin{pmatrix} B_1 Q_1 & & * \\ & \ddots & \\ 0 & & B_r Q_r \end{pmatrix}$$

$$= \begin{pmatrix} Q_1^* B_1 Q_1 & & * \\ & \ddots & \\ 0 & & Q_r^* B_r Q_r \end{pmatrix} = \begin{pmatrix} T_1 & & * \\ & \ddots & \\ 0 & & T_r \end{pmatrix}$$

Now this is a real upper triangular matrix and the eigenvalues of A consist of the union of the eigenvalues of the T_i which is the same as the union of the eigenvalues of the B_i. ■

Here is the description of the great and glorious QR algorithm.

The QR Algorithm

Let A be an $n \times n$ real matrix. Let $A_0 = A$. Suppose that A_{k-1} has been found. To find A_k let

$$A_{k-1} = Q_k R_k,\ A_k = R_k Q_k,$$

where $Q_k R_k$ is a QR factorization of A_{k-1}. Thus R is upper triangular with nonnegative entries on the main diagonal and Q is real and unitary (orthogonal).

15.6 MATLAB AND THE QR ALGORITHM

This is most easily done in MATLAB. Given H you would then just do the QR algorithm on this matrix to get eigenvalues. The syntax for doing this is as follows. Here 50 iterations are being used.

```
H=[enter H here]
hold on
```

```
for k=1:50
[Q,R]=qr(H);
H=R*Q;
end
Q
R
H
```

Of course if MATLAB already knows H then you don't need to re-enter it. This happens when you use MATLAB to find an upper Hessenberg matrix similar to the original matrix. This is discussed later.

The main significance of this algorithm is in the following easy but important theorem.

Theorem 15.6.1 *Let A be any $n \times n$ complex matrix and let $\{A_k\}$ be the sequence of matrices described above by the QR algorithm. Then each of these matrices is unitarily similar to A.*

Proof: Clearly A_0 is orthogonally similar to A because they are the same thing. Suppose then that

$$A_{k-1} = Q^* A Q$$

Then from the algorithm,

$$A_{k-1} = Q_k R_k, \; R_k = Q_k^* A_{k-1}$$

Therefore, from the algorithm,

$$A_k \equiv R_k Q_k = Q_k^* A_{k-1} Q_k = Q_k^* Q^* A Q Q_k = (Q Q_k)^* \, A Q Q_k,$$

and so A_k is unitarily similar to A also. ■

Although the sequence $\{A_k\}$ may fail to converge, it is nevertheless often the case that for large k, A_k is of the form

$$A_k = \begin{pmatrix} B_k & & * \\ & \ddots & \\ e & & B_r \end{pmatrix}$$

where the B_i are blocks which run down the diagonal of the matrix, and all of the entries below this block diagonal are very small. Then letting T_B denote the matrix obtained by setting all of these small entries equal to zero, one can argue, using methods of analysis, that the eigenvalues of A_k are close to the eigenvalues of T_B. From Lemma 15.5.1 the eigenvalues of T_B are the eigenvalues of the blocks B_i. Thus, the eigenvalues of A are the same as those of A_k and these are close to the eigenvalues of T_B.

In proving things about this algorithm and also for the sake of convenience, here is a technical result.

Corollary 15.6.2 *For Q_k, R_k, A_k given in the QR algorithm,*

$$A = Q_1 \cdots Q_k A_k Q_k^* \cdots Q_1^* \tag{15.3}$$

For $Q^{(k)} \equiv Q_1 \cdots Q_k$ and $R^{(k)} \equiv R_k \cdots R_1$, it follows that

$$A^k = Q^{(k)} R^{(k)}$$

Here A_k is the usual thing, A raised to the k^{th} power.

Proof: From the algorithm,

$$A = A_0 = Q_1 R_1,\ Q_1^* A_0 = R_1,\ A_1 \equiv R_1 Q_1 = Q_1^* A Q_1$$

Hence

$$Q_1 A_1 Q_1^* = A$$

Suppose the formula 15.3 holds for k. Then from the algorithm,

$$A_k = Q_{k+1} R_{k+1},\ R_{k+1} = Q_{k+1}^* A_k,\ A_{k+1} = R_{k+1} Q_{k+1} = Q_{k+1}^* A_k Q_{k+1}$$

Hence $Q_{k+1} A_{k+1} Q_{k+1}^* = A_k$ and so

$$A = Q_1 \cdots Q_k A_k Q_k^* \cdots Q_1^* = Q_1 \cdots Q_k Q_{k+1} A_{k+1} Q_{k+1}^* Q_k^* \cdots Q_1^*$$

This shows the first part.

The second part is clearly true from the algorithm if $k = 1$. Then from the first part and the algorithm,

$$A = Q_1 \cdots Q_k Q_{k+1} A_{k+1} Q_{k+1}^* Q_k^* \cdots Q_1^* = Q_1 \cdots Q_k Q_{k+1} R_{k+1} \overbrace{Q_{k+1} Q_{k+1}^*}^{I} Q_k^* \cdots Q_1^*$$

It follows that

$$\begin{aligned} A^{k+1} &= A A^k = Q_1 \cdots Q_k Q_{k+1} R_{k+1} Q_k^* \cdots Q_1^* Q^{(k)} R^{(k)} \\ &= Q^{(k+1)} R_{k+1} \left(Q^{(k)}\right)^* Q^{(k)} R^{(k)} \end{aligned}$$

Hence

$$A^{k+1} = Q^{(k+1)} R_{k+1} R^{(k)} = Q^{(k+1)} R^{(k+1)} \blacksquare$$

Now suppose that A^{-1} exists. How do two QR factorizations compare? Since A^{-1} exists, it would require that if $A = QR$, then R^{-1} must exist. Now an upper triangular matrix has inverse which is also upper triangular. This follows right away from the algorithm presented early in the book for finding the inverse. If $A = Q_1 R_1 = Q_2 R_2$, then $Q_1^* Q_2 = R_1 R_2^{-1}$ and so $R_1 R_2^{-1}$ is an upper triangular matrix which is also unitary and in addition has all positive entries down the diagonal. For simplicity, call it R. Thus R is upper triangular and $RR^* = R^* R = I$. It follows easily that R must equal I and so $R_1 = R_2$ which requires $Q_1 = Q_2$.

Now in the above corollary, you know that

$$A = Q_1 \cdots Q_k A_k Q_k^* \cdots Q_1^* = Q^{(k)} A_k \left(Q^{(k)}\right)^*$$

Also, from this corollary, you know that

$$A^k = Q^{(k)} R^{(k)}$$

You could also simply take the QR factorization of A^k to obtain $A^k = QR$. Then from what was just pointed out, if exists,

$$Q^{(k)} = Q$$

Thus from the above corollary,

$$A_k = \left(Q^{(k)}\right)^* A Q^{(k)} = Q^* A Q$$

Therefore, in using the QR algorithm in the case where A has an inverse, it suffices to take

$$A^k = QR$$

and then consider the matrix

$$Q^* A Q = A_k.$$

This is so theoretically. In practice it might not work out all that well because of round off errors.

There is also an interesting relation to the power method. Let

$$A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$$

Then from the way we multiply matrices,

$$A^{k+1} = \begin{pmatrix} A^k \mathbf{a}_1 & \cdots & A^k \mathbf{a}_n \end{pmatrix}$$

and for large k, $A^k \mathbf{a}_i$ would be expected to point roughly in the direction of the eigenvector corresponding to the largest eigenvalue. Then if you form the QR factorization,

$$A^{k+1} = QR$$

the columns of Q are an orthonormal basis obtained essentially from the Gram Schmidt procedure. Thus the first column of Q has roughly the direction of an eigenvector associated with the largest eigenvalue of A. It follows that the first column of Q^*AQ is approximately equal to $\lambda_1 \mathbf{q}_1$ and so the top entry will be close to $\lambda_1 \mathbf{q}_1^* \mathbf{q}_1 = \lambda_1$ and the entries below it are close to 0. Thus the eigenvalues of the matrix should be close to this top entry of the first column along with the eigenvalues of the $(n-1) \times (n-1)$ matrix in the lower right corner. If this is a 2×2 you can find the eigenvalues using the quadratic formula. If it is larger, you could just use the same procedure for finding its eigenvalues but now you are dealing with a smaller matrix.

15.6.1 THE UPPER HESSENBERG FORM

Actually, when using the QR algorithm, contrary to what is discussed above, you should always deal with a matrix which is similar to the given matrix which is in upper Hessenberg form. This means all the entries below the sub diagonal equal 0. Here is an easy lemma.

Lemma 15.6.3 *Let A be an $n \times n$ matrix. Then it is unitarily similar to a matrix in upper Hessenberg form and this similarity can be computed.*

Proof: Let A be an $n \times n$ matrix. Suppose $n > 2$. There is nothing to show otherwise.

$$A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{d} & A_1 \end{pmatrix}$$

where A_1 is $n-1 \times n-1$. Consider the $n-1 \times 1$ matrix $\mathbf{d}$. Then let Q be a Householder reflection such that

$$Q\mathbf{b} = \begin{pmatrix} c \\ \mathbf{0} \end{pmatrix} \equiv \mathbf{c}$$

Then

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q \end{pmatrix} \begin{pmatrix} a & \mathbf{b} \\ \mathbf{d} & A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q^* \end{pmatrix} = \begin{pmatrix} a & \mathbf{b}Q^* \\ \mathbf{c} & QA_1Q^* \end{pmatrix}$$

By similar reasoning, there exists an $n-1 \times n-1$ matrix

$$U = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{pmatrix}$$

such that

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{pmatrix} QA_1Q^* \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1^* \end{pmatrix} = \begin{pmatrix} * & * & \cdots & * \\ * & * & \ddots & \vdots \\ & \ddots & \ddots & * \\ \mathbf{0} & & * & * \end{pmatrix}$$

Thus

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{pmatrix} \begin{pmatrix} a & \mathbf{b}Q^* \\ \mathbf{c} & QA_1Q^* \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U^* \end{pmatrix}$$

will have all zeros below the first two entries on the sub diagonal. Continuing this way shows the result. ■

Not surprisingly, MATLAB can find a Hessenberg matrix for a given square matrix. Here is the syntax.

```
A=[2 1 3;-5,3,-2;1,2,3];
[P,H]=hess(A)
P'*A*P
```

The output will be first P a unitary matrix, then H, a Hessenberg matrix and finally, the last line just verifies that $P^*AP = H$. Then you can do the QR algorithm on H.

The reason you should use a matrix which is upper Hessenberg and similar to A in the QR algorithm is that the algorithm keeps returning a matrix in upper Hessenberg form and if you are looking for block upper triangular matrices, this will force the size of the blocks to be no larger than 2×2 which are easy to handle using the quadratic formula. This is in the following lemma.

Lemma 15.6.4 *Let $\{A_k\}$ be the sequence of iterates from the QR algorithm, A^{-1} exists. Then if A_k is upper Hessenberg, so is A_{k+1}.*

Proof: The matrix is upper Hessenberg means that $A_{ij} = 0$ whenever $i - j \geq 2$.

$$A_{k+1} = R_k Q_k$$

where $A_k = Q_k R_k$. Therefore $A_k R_k^{-1} = Q_k$ and so

$$A_{k+1} = R_k Q_k = R_k A_k R_k^{-1}$$

Let the ij^{th} entry of A_k be a_{ij}^k. Then if $i - j \geq 2$

$$a_{ij}^{k+1} = \sum_{p=i}^{n} \sum_{q=1}^{j} r_{ip} a_{pq}^k r_{qj}^{-1}$$

It is given that $a_{pq}^k = 0$ whenever $p - q \geq 2$. However, from the above sum,

$$p - q \geq i - j \geq 2,$$

and so the sum equals 0. ■

Example 15.6.5 *Find the solutions to the equation $x^4 - 4x^3 + 8x^2 - 8x + 4 = 0$ using the QR algorithm.*

This is the characteristic equation of the matrix

$$H = \begin{pmatrix} 4 & -8 & 8 & -4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Since the constant term in the equation is not 0, it follows that the matrix has an inverse. It is already in upper Hessenberg form. Lets apply the QR algorithm above with 100 iterations. This yields the following matrix which is similar to H.

$$\begin{pmatrix} .6761 & -.7372 & .0469 & .1593 \\ 1.5344 & 1.3435 & 3.1593 & 10.7276 \\ 0 & 0.0001 & -1.3435 & -4.1885 \\ 0 & 0 & 1.5344 & 3.3239 \end{pmatrix}$$

The number in the third row and second column is so small that we neglect it. All that remains is to find the eigenvalues are the two blocks

$$\begin{pmatrix} .6761 & -.7372 \\ 1.5344 & 1.3435 \end{pmatrix}, \begin{pmatrix} -1.3435 & -4.1885 \\ 1.5344 & 3.3239 \end{pmatrix}$$

Thus the eigenvalues of the original matrix are those which result from these two blocks. Since these are 2×2 matrices, you can find the answer from the quadratic formula. Thus the eigenvalues are

$$\begin{aligned} &1.\,009\,8 + 1.\,009\,9i, 1.\,009\,8 - 1.\,009\,9i, \\ &0.990\,2 + 0.990\,29i, 0.990\,2 - 0.990\,29i \end{aligned}$$

In fact, the eigenvalues are exactly $1+i, 1+i, 1-i, 1-i$ listed according to multiplicity. Now you can use the shifted inverse power method to find much better approximations as well as eigenvectors. For example, you would use that algorithm to find the eigenvalue close to $0.990\,2 + 0.990\,29i$ along with the associated eigenvector. It yields $1+i$ as the eigenvalue closest to $0.99 + 0.99i$ along with the eigenvector

$$\begin{pmatrix} 1 & 0.5-0.5i & -0.5i & -0.25-0.25i \end{pmatrix}^T$$

Of course we didn't care about the eigenvector but there it is anyway. It took 844 iterations even though $0.99 + 0.99i$ was very close to the true eigenvalue.

Example 15.6.6 *Find the eigenvalues for the symmetric matrix*

$$A = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 2 & 0 & 1 & 3 \\ 3 & 1 & 3 & 2 \\ -1 & 3 & 2 & 1 \end{pmatrix}$$

Also find an eigenvector.

A Hessenberg matrix similar to the above matrix is

$$H = \begin{pmatrix} .0888 & -.6421 & 0 & 0 \\ -.6421 & 3.8398 & -3.348 & 0 \\ 0 & -3.348 & .0714 & -3.7417 \\ 0 & 0 & -3.7417 & 1 \end{pmatrix}$$

Thus you could use the QR algorithm on this to identify the eigenvalues. In using this algorithm, MATLAB already knows H unless you did clear all or close all. Thus you don't need to enter the matrix in the QR algorithm. You just need to refer to H.

This yields

$$\begin{pmatrix} 6.643 & 0 & 0 & 0 \\ 0 & -4.1018 & 0 & 0 \\ 0 & 0 & 2.4589 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus the eigenvalues are those diagonal entries. Now if you want to find eigenvectors, there is a way to keep track of things and get them from the above, but you could also simply go back to the matrix A and use the shifted inverse power method. Starting with one of these approximate eigenvalues or a number close to one as α. I shall pick the eigenvalue 6.643 and obtain an eigenvector and possibly a better approximation to this eigenvalue using the shifted inverse power method using the iterative procedure given above. This yields the eigenvector

$$\mathbf{u} = \begin{pmatrix} 0.6442 & 0.5961 & 1 & 0.5572 \end{pmatrix}^T$$

which works extremely well, along with the eigenvalue 6.643. In fact, the error between $A\mathbf{u}$ and $6.643\mathbf{u}$ is on the order of 10^{-14}.

15.7 EXERCISES

1. Using the power method, find the eigenvalue correct to one decimal place having largest absolute value for the matrix $A = \begin{pmatrix} 0 & -4 & -4 \\ 7 & 10 & 5 \\ -2 & 0 & 6 \end{pmatrix}$ along with an eigenvector associated with this eigenvalue.
2. Using the power method, find the eigenvalue correct to one decimal place having largest absolute value for the matrix $A = \begin{pmatrix} 15 & 6 & 1 \\ -5 & 2 & 1 \\ 1 & 2 & 7 \end{pmatrix}$ along with an eigenvector associated with this eigenvalue.
3. Using the power method, find the eigenvalue correct to one decimal place having largest absolute value for the matrix $A = \begin{pmatrix} 10 & 4 & 2 \\ -3 & 2 & -1 \\ 0 & 0 & 4 \end{pmatrix}$ along with an eigenvector associated with this eigenvalue.
4. Using the power method, find the eigenvalue correct to one decimal place having largest absolute value for the matrix $A = \begin{pmatrix} 15 & 14 & -3 \\ -13 & -18 & 9 \\ 5 & 10 & -1 \end{pmatrix}$ along with an eigenvector associated with this eigenvalue.
5. In Example 15.4.3 an eigenvalue was found correct to several decimal places along with an eigenvector. Find the other eigenvalues along with their eigenvectors.
6. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically. In this case the exact eigenvalues are $\pm\sqrt{3}, 6$. Compare with the exact answers.

7. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

   The exact eigenvalues are $2, 4+\sqrt{15}, 4-\sqrt{15}$. Compare your numerical results with the exact values. Is it much fun to compute the exact eigenvectors?

8. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

   We don't know the exact eigenvalues in this case. Check your answers by multiplying your numerically computed eigenvectors by the matrix.

9. Find the eigenvalues and eigenvectors of the matrix $A = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ numerically.

   We don't know the exact eigenvalues in this case. Check your answers by multiplying your numerically computed eigenvectors by the matrix.

10. Consider the matrix $A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 0 \end{pmatrix}$ and the vector $(1, 1, 1)^T$. Estimate the distance between the Rayleigh quotient determined by this vector and some eigenvalue of A.

11. Consider the matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 4 \\ 1 & 4 & 5 \end{pmatrix}$ and the vector $(1, 1, 1)^T$. Estimate the distance between the Rayleigh quotient determined by this vector and some eigenvalue of A.

12. Using Gerschgorin's theorem, find upper and lower bounds for the eigenvalues of

$$A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 6 & 4 \\ 3 & 4 & -3 \end{pmatrix}.$$

13. The QR algorithm works very well on general matrices. Try QR the algorithm on the following matrix which happens to have some complex eigenvalues.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

Use the QR algorithm to get approximate eigenvalues and then use the shifted inverse power method on one of these to get an approximate eigenvector for one of the complex eigenvalues.

14. Use the QR algorithm to approximate the eigenvalues of the symmetric matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & -8 & 1 \\ 3 & 1 & 0 \end{pmatrix}$$

15. Try to find the eigenvalues of the matrix $\begin{pmatrix} 3 & 3 & 1 \\ -2 & -2 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ using the QR algorithm. It has eigenvalues $1, i, -i$. You will see the algorithm won't work well. ►

16. Let $q(\lambda) = a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$. Now consider the **companion matrix,**

$$C \equiv \begin{pmatrix} 0 & \cdots & 0 & -a_0 \\ 1 & 0 & & -a_1 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & -a_{n-1} \end{pmatrix}$$

Show that $q(\lambda)$ is the characteristic equation for C. Thus the roots of $q(\lambda)$ are the eigenvalues of C. You can prove something similar for

$$C = \begin{pmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_0 \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{pmatrix}$$

Hint: The characteristic equation is

$$\det \begin{pmatrix} \lambda & \cdots & 0 & a_0 \\ -1 & \lambda & & a_1 \\ & \ddots & \ddots & \vdots \\ 0 & & -1 & \lambda + a_{n-1} \end{pmatrix}$$

Expand along the first column. Thus

$$\lambda \begin{vmatrix} \lambda & \cdots & 0 & a_1 \\ -1 & \lambda & & a_2 \\ & \ddots & \ddots & \vdots \\ 0 & & -1 & \lambda + a_{n-1} \end{vmatrix} + \begin{vmatrix} 0 & 0 & \cdots & a_0 \\ -1 & \lambda & \cdots & a_2 \\ \vdots & \ddots & & \vdots \\ 0 & & -1 & \lambda + a_3 \end{vmatrix}$$

Now use induction on the first term and for the second, note that you can expand along the top row to get

$$(-1)^{n-2} a_0 (-1)^n = a_0.$$

17. Suppose A is a real symmetric, invertible, matrix, or more generally one which has real eigenvalues. Then as described above, it is typically the case that

$$A^p = Q_1 R$$

and

$$Q_1^T A Q_1 = \begin{pmatrix} a_1 & \mathbf{b}_1^T \\ \mathbf{e}_1 & A_1 \end{pmatrix}$$

where $\mathbf{e}$ is very small. Then you can do the same thing with A_1 to obtain another smaller orthogonal matrix Q_2 such that

$$Q_2^T A_1 Q_2 = \begin{pmatrix} a_2 & \mathbf{b}_2^T \\ \mathbf{e}_2 & A_2 \end{pmatrix}$$

Explain why

$$\begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & Q_2 \end{pmatrix}^T Q_1^T A Q_1 \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & Q_2 \end{pmatrix} = \begin{pmatrix} a_1 & & * \\ \vdots & a_2 & \\ \mathbf{e}_1 & \mathbf{e}_2 & A_3 \end{pmatrix}$$

where the $\mathbf{e}_i$ are very small. Explain why one can construct an orthogonal matrix Q such that

$$Q^T A Q = (T + E)$$

where T is an upper triangular matrix and E is very small. In case A is symmetric, explain why T is actually a diagonal matrix. Next explain why, in the case of A symmetric, that the columns of Q are an orthonormal basis of vectors, each of which is close to an eigenvector. Thus this will compute, not just the eigenvalues but also the eigenvectors.

18. Explain how one could use the QR algorithm or the above procedure to compute the singular value decomposition of an arbitrary real $m \times n$ matrix.

BIBLIOGRAPHY

[1] **Apostol T.** *Calculus Volume II Second edition, Wiley 1969.*

[2] **Baker, Roger,** *Linear Algebra, Rinton Press 2001.*

[3] **Davis H. and Snider A.,** *Vector Analysis Wm. C. Brown 1995.*

[4] **Edwards C.H**. *Advanced Calculus of several Variables, Dover 1994.*

[5] **Chahal J.S.,** *Historical Perspective of Mathematics 2000 B.C. - 2000 A.D. Kendrick Press, Inc. (2007)*

[6] **Golub, G. and Van Loan, C.,***Matrix Computations, Johns Hopkins University Press, 1996.*

[7] **Greenberg M.D.** *Advanced Engineering Mathematics Prentice Hall 1998 Second edition.*

[8] **Gurtin M.** *An introduction to continuum mechanics, Academic press 1981.*

[9] **Hardy G.** *A Course Of Pure Mathematics, Tenth edition, Cambridge University Press 1992.*

[10] **Horn R. and Johnson C.** *matrix Analysis, Cambridge University Press, 1985.*

[11] **Jacobsen N.** *Basic Algebra Freeman 1974.*

[12] **Karlin S. and Taylor H.** *A First Course in Stochastic Processes, Academic Press, 1975.*

[13] **Kuttler K.***Linear Algebra On web page.* Linear Algebra

[14] **Nobel B. and Daniel J.** *Applied Linear Algebra, Prentice Hall, 1977.*

[15] **Rudin W.** *Principles of Mathematical Analysis, McGraw Hill, 1976.*

[16] **Salas S. and Hille E.,** *Calculus One and Several Variables, Wiley 1990.*

[17] **Strang Gilbert,** *Linear Algebra and its Applications, Harcourt Brace Jovanovich 1980.*

[18] **Wilkinson, J.H.,** *The Algebraic Eigenvalue Problem, Clarendon Press Oxford 1965.*

[19] Yosida K., *Functional Analysis, Springer Verlag, 1978.*

APPENDIX C ANSWERS TO SELECTED EXERCISES

C.10 EXERCISES 319

1. $$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & -3 & 3 \\ 0 & 0 & 3 \end{pmatrix}$$

3. $$\begin{pmatrix} 1 & -2 & -5 & 0 \\ -2 & 5 & 11 & 3 \\ 3 & -6 & -15 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -5 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

5. $$\begin{pmatrix} 1 & -3 & -4 & -3 \\ -3 & 10 & 10 & 10 \\ 1 & -6 & 2 & -5 \end{pmatrix} =$$
$$\begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3 & -4 & -3 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

7. $$\begin{pmatrix} 3 & -2 & 1 \\ 9 & -8 & 6 \\ -6 & 2 & 2 \\ 3 & 2 & -7 \end{pmatrix} =$$
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & 1 & 1 & 0 \\ 1 & -2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 3 & -2 & 1 \\ 0 & -2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

9. $$\begin{pmatrix} -1 & -3 & -1 \\ 1 & 3 & 0 \\ 3 & 9 & 0 \\ 4 & 12 & 16 \end{pmatrix} =$$
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ -4 & 0 & -4 & 1 \end{pmatrix} \begin{pmatrix} -1 & -3 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix}$$

11. An LU factorization of the coefficient matrix is

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 2 & 3 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

First solve

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$
$$= \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}$$

which yields $u = 1, v = 2, w = 6$. Next solve

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}$$

This yields $z = 6, y = -16, x = 27$.

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

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

15. $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \end{pmatrix} =$

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \cdot$

$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & 1 \end{pmatrix}$

17. $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 2 & 4 & 1 \\ 3 & 2 & 1 \end{pmatrix} =$

$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix} \cdot$

$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -4 & -2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$

20. $\begin{pmatrix} \frac{1}{11}\sqrt{11} & \frac{13}{66}\sqrt{2}\sqrt{11} & -\frac{1}{6}\sqrt{2} \\ \frac{3}{11}\sqrt{11} & -\frac{5}{66}\sqrt{2}\sqrt{11} & -\frac{1}{6}\sqrt{2} \\ \frac{1}{11}\sqrt{11} & \frac{1}{33}\sqrt{2}\sqrt{11} & \frac{2}{3}\sqrt{2} \end{pmatrix} \cdot$

$\begin{pmatrix} \sqrt{11} & -\frac{4}{11}\sqrt{11} & \frac{6}{11}\sqrt{11} \\ 0 & \frac{6}{11}\sqrt{2}\sqrt{11} & \frac{2}{11}\sqrt{2}\sqrt{11} \\ 0 & 0 & \sqrt{2} \end{pmatrix}$

22. You would have $QR\mathbf{x} = \mathbf{b}$ and so then you would have $R\mathbf{x} = Q^T\mathbf{b}$. Now R is upper triangular and so the solution of this problem is fairly simple.

C.11 EXERCISES 352

1. The minimum is $-11/2$ and it occurs when $x_1 = x_3 = x_6 = 0$ and
   $x_2 = 7/2, x_4 = 13/2, x_6 = -11/2$
   The maximum is 7 and it occurs when $x_1 = 7, x_2 = 0, x_3 = 0, x_4 = 3, x_5 = 5, x_6 = 0.$

2. Maximize and minimize the following if possible. All variables are nonnegative.

   a) The maximum is 7 when $x_1 = 7$ and $x_1, x_3 = 0.$
      The maximum is -7 and it happens when $x_1 = 0, x_2 = 7/2, x_3 = 0.$
   b) The maximum is -21 and it occurs when $x_1 = x_2 = 0, x_3 = 7.$
      The maximum is 7 and it occurs when $x_1 = 7, x_2 = 0, x_3 = 0.$
   c) The maximum is 0 and it occurs when $x_1 = x_2 = 0, x_3 = 1.$
      The maximum is 14 and it happens when $x_1 = 7, x_2 = x_3 = 0.$
   d) The maximum is 7 and it happens when $x_2 = 7/2, x_3 = x_1 = 0.$
      The maximum is 0 when $x_1 = x_2 = 0, x_3 = 1.$

4. Find solutions if possible.

   a) There is no solution to these inequalities with $x_1, x_2 \geq 0.$
   b) A solution is $x_1 = 8/5, x_2 = x_3 = 0.$
   c) No solution to these inequalities for which all the variables are nonnegative.
   d) There is a solution when $x_2 = 2, x_3 = 0, x_1 = 0.$
   e) There is no solution.

C.12 EXERCISES 396

4. If it did have $\lambda \in \mathbb{R}$ as an eigenvalue, then there would exist a vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for λ a real number. Therefore, $A\mathbf{x}$ and $\mathbf{x}$ would need to be parallel. However, this doesn't happen because A rotates the vectors.
6. $A^m\mathbf{x} = \lambda^m\mathbf{x}$ any integer. In the case of -1,

   $$A^{-1}\lambda\mathbf{x} = AA^{-1}\mathbf{x} = \mathbf{x}$$

   so $A^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}$. Thus the eigenvalues of A^{-1} are just λ^{-1} where λ is an eigenvalue of A.

7. Let x be the eigenvector. Then $A^m\mathbf{x} = \lambda^m\mathbf{x}, A^m\mathbf{x} = A\mathbf{x} = \lambda\mathbf{x}$ and so

   $$\lambda^m = \lambda$$

Hence if $\lambda \neq 0$, then

$$\lambda^{m-1} = 1$$

and so $|\lambda| = 1$.

10. eigenvectors: $\left\{\begin{matrix} 3 \\ 1 \\ 1 \end{matrix}\right\} \leftrightarrow 1, \left\{\begin{matrix} 2 \\ 1 \\ 1 \end{matrix}\right\} \leftrightarrow 2.$ This is a defective matrix.

11. eigenvectors: $\left\{\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}\right\} \leftrightarrow -1,$

$$\left\{\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\right\} \leftrightarrow 2$$

This matrix is not defective because, even though $\lambda = 1$ is a repeated eigenvalue, it has a 2 dimensional eigenspace.

14. eigenvectors: $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{1}{2} \\ 1 \end{pmatrix} \right\} \leftrightarrow 3,$

$\left\{ \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \right\} \leftrightarrow 6$

This matrix is not defective.

16. eigenvectors: $\left\{ \begin{pmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5}{6} \\ 0 \\ 1 \end{pmatrix} \right\} \leftrightarrow -1$

This matrix is defective. In this case, there is only one eigenvalue, -1 of multiplicity 3 but the dimension of the eigenspace is only 2.

19. eigenvectors: $\left\{ \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ 1 \end{pmatrix} \right\} \leftrightarrow 0$ This one is defective.

20. eigenvectors: $\left\{ \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ 1 \end{pmatrix} \right\} \leftrightarrow 1$

This is defective.

22. eigenvectors:

$\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\} \leftrightarrow 4, \left\{ \begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} \right\}$

$\leftrightarrow 2-2i, \left\{ \begin{pmatrix} i \\ i \\ 1 \end{pmatrix} \right\} \leftrightarrow 2+2i$

24. eigenvectors:

$\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\} \leftrightarrow 4,$

$\left\{ \begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} \right\} \leftrightarrow 2-2i,$

$\left\{ \begin{pmatrix} i \\ i \\ 1 \end{pmatrix} \right\} \leftrightarrow 2+2i$ This matrix is not defective.

26. eigenvectors:

$$\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\} \leftrightarrow -6,$$
$$\left\{\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix}\right\} \leftrightarrow 2-6i,$$
$$\left\{\begin{pmatrix} i \\ i \\ 1 \end{pmatrix}\right\} \leftrightarrow 2+6i$$

28. The characteristic polynomial is of degree three and it has real coefficients. Therefore, there is a real root and two distinct complex roots. It follows that A cannot be defective because it has three distinct eigenvalues.

29. eigenvectors: $\left\{\begin{pmatrix} -i \\ 1 \end{pmatrix}\right\} \leftrightarrow -i,$

$$\left\{\begin{pmatrix} i \\ 1 \end{pmatrix}\right\} \leftrightarrow i$$

32. eigenvectors: $\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\} \leftrightarrow -1,$

$$\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\} \leftrightarrow 1$$

34. eigenvectors: $\left\{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\} \leftrightarrow 1,$

$$\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\} \leftrightarrow 2,$$
$$\left\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}\right\} \leftrightarrow 3$$

36. In terms of percentages in the various locations,

$$\begin{pmatrix} 21.429 \\ 21.429 \\ 57.143 \end{pmatrix}$$

38. Obviously A cannot be onto because the range of A has dimension 1 and the dimension of this space should be 3 if the matrix is onto. Therefore, A cannot be invertible. Its row reduced echelon form cannot be I since if it were, A would be onto. $A\mathbf{w} = \mathbf{w}$ so it has an eigenvalue equal to 1. Now suppose $A\mathbf{x} = \lambda\mathbf{x}$. Thus, from the Cauchy Schwarz inequality,

$$\begin{aligned} |\mathbf{x}| &= \frac{|\mathbf{x}|\,|\mathbf{w}|}{|\mathbf{w}|^2}\,|\mathbf{w}| \\ &\geq \frac{|(\mathbf{x},\mathbf{w})|}{|\mathbf{w}|^2}\,|\mathbf{w}| = |\lambda|\,|\mathbf{x}| \end{aligned}$$

and so $|\lambda| \leq 1$.

40. Since the vectors are linearly independent, the matrix S has an inverse. Denoting this inverse by

$$S^{-1} = \begin{pmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_n^T \end{pmatrix}$$

it follows by denition that

$$\mathbf{w}_i^T \mathbf{x}_j = \delta_{ij}.$$

Therefore,

$$S^{-1}MS = S^{-1}(M\mathbf{x}_1, \cdots, M\mathbf{x}_n)$$
$$= \begin{pmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_n^T \end{pmatrix} (\lambda_1 \mathbf{x}_1, \cdots, \lambda_n \mathbf{x}_n)$$

$$= \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

42. The diagonally dominant condition implies that none of the Gerschgorin disks contain 0. Therefore, 0 is not an eigenvalue. Hence A is one to one, hence invertible.

44. First note that $(AB)^* = B^*A^*$. Say $M\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$. Then

$$\begin{aligned} \overline{\lambda} |\mathbf{x}|^2 &= \overline{\lambda}\mathbf{x}^*\mathbf{x} = (\lambda\mathbf{x})^* \mathbf{x} \\ &= (M\mathbf{x})^* \mathbf{x} = \mathbf{x}^* M^* \mathbf{x} \end{aligned}$$

$$= \mathbf{x}^* M\mathbf{x} = \mathbf{x}^* \lambda\mathbf{x} = \lambda |\mathbf{x}|^2$$

Hence $\lambda = \overline{\lambda}$.

47. Suppose A is skew symmetric. Then what about iA?

$$(iA)^* = -iA^* = -iA^T = iA$$

and so iA is self adjoint. Hence it has all real eigenvalues. Therefore, the eigenvalues of A are all of the form $i\lambda$ where λ is real. Now what about the eigenvectors? You need

$$A\mathbf{x} = i\lambda\mathbf{x}$$

where $\lambda \neq 0$ is real and A is real. Then

$$A\mathrm{Re}(\mathbf{x}) = i\lambda\mathrm{Re}(\mathbf{x})\text{'}$$

The left has all real entries and the right has all pure imaginary entries. Hence $\mathrm{Re}(\mathbf{x}) = \mathbf{0}$ and so x has all imaginary entries.

C.13 EXERCISES 449

1. a. orthogonal and transformation, b. symmetric, c. skew symmetric.

4. $$\begin{aligned}&\|U\mathbf{x}\|^2 = (U\mathbf{x}, U\mathbf{x})\\ &= \left(U^T U\mathbf{x}, \mathbf{x}\right) = (I\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|^2\end{aligned}$$

   Next suppose distance is preserved by U. Then

$$\begin{aligned} & (U(\mathbf{x}+\mathbf{y}), U(\mathbf{x}+\mathbf{y})) \\ = & \|Ux\|^2 + \|Uy\|^2 + 2(Ux, Uy) \\ = & \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\left(U^T Ux, \mathbf{y}\right) \end{aligned}$$

   But since U preserves distances, it is also the case that

$$\begin{aligned} & (U(\mathbf{x}+\mathbf{y}), U(\mathbf{x}+\mathbf{y})) \\ = & \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2(\mathbf{x}, \mathbf{y}) \end{aligned}$$

   Hence

$$(\mathbf{x}, \mathbf{y}) = \left(U^T Ux, \mathbf{y}\right)$$

   and so

$$\left(\left(U^T U - I\right)\mathbf{x}, \mathbf{y}\right) = 0$$

   Since y is arbitrary, it follows that $U^T U - I = 0$. Thus U is orthogonal.

5. $$\begin{pmatrix} x & y & z \end{pmatrix} \cdot \begin{pmatrix} a_1 & a_4/2 & a_5/2 \\ a_4/2 & a_2 & a_6/2 \\ a_5/2 & a_6/2 & a_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

7. If A is symmetric, then $A = U^T DU$ for some D a diagonal matrix in which all the diagonal entries are non zero. Hence $A^{-1} = U^{-1}D^{-1}U^{-T}$. Now

$$\begin{aligned} U^{-1}U^{-T} &= \left(U^T U\right)^{-1} \\ &= I^{-1} = I \end{aligned}$$

   and so $A^{-1} = QD^{-1}Q^T$, where Q is orthogonal. Is this thing on the right symmetric? Take its transpose. This is $QD^{-1}Q^T$ which is the same thing, so it appears that a symmetric matrix must have symmetric inverse. Now consider raising it to a power.

$$A^k = U^T D^k U$$

and the right side is clearly symmetric.

8. Yes.

11. eigenvectors: $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\} \leftrightarrow c,$

$$\left\{\begin{pmatrix}0\\-i\\1\end{pmatrix}\right\} \leftrightarrow -ib,$$
$$\left\{\begin{pmatrix}0\\i\\1\end{pmatrix}\right\} \leftrightarrow ib$$

12. eigenvectors: $\left\{\begin{pmatrix}0\\-i\\1\end{pmatrix}\right\} \leftrightarrow a-ib,$

$$\left\{\begin{pmatrix}0\\i\\1\end{pmatrix}\right\} \leftrightarrow a+ib,$$
$$\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\} \leftrightarrow c$$

13. eigenvectors:

$$\left\{\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}\right\} \leftrightarrow 6,$$
$$\left\{\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}\right\} \leftrightarrow 12,$$

$$\left\{\frac{1}{\sqrt{6}}\begin{pmatrix}-1\\-1\\2\end{pmatrix}\right\} \leftrightarrow 18$$

15.
$$\begin{aligned}\lambda\mathbf{x}^T\overline{\mathbf{x}} &= (A\mathbf{x})^T\,\overline{\mathbf{x}}\\ &\overset{(CD)^T=D^TC^T}{=} \mathbf{x}^T A\overline{\mathbf{x}}\\ &\overset{A\text{ is real}}{=} \mathbf{x}^T\overline{A\mathbf{x}}\\ &\overset{\lambda\text{ is eigenvalue}}{=} \mathbf{x}^T\overline{\lambda}\overline{\mathbf{x}} = \overline{\lambda}\mathbf{x}^T\overline{\mathbf{x}}\end{aligned}$$

and so $\lambda = \overline{\lambda}$. This shows that all eigenvalues are real. It follows all the eigenvectors are real. Why?

Because A is real. $A\mathbf{x} = \lambda\mathbf{x}$, $A\overline{\mathbf{x}} = \lambda\overline{\mathbf{x}}$, so $\mathbf{x}+\overline{\mathbf{x}}$ is an eigenvector. Hence it can be assumed all eigenvectors are real.

Now let $\mathbf{x}, \mathbf{y}, \mu$ and λ be given as above.

$$\begin{aligned} \lambda(\mathbf{x}\cdot\mathbf{y}) &= \lambda\mathbf{x}\cdot\mathbf{y} \\ &= A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A\mathbf{y} \\ &= \mathbf{x}\cdot\mu\mathbf{y} = \mu(\mathbf{x}\cdot\mathbf{y}) \\ &= \mu(\mathbf{x}\cdot\mathbf{y}) \end{aligned}$$

and so

$$(\lambda - \mu)\,\mathbf{x}\cdot\mathbf{y} = 0.$$

Since $\lambda \neq \mu$, it follows $\mathbf{x}\cdot\mathbf{y} = 0$.

17.
$$\begin{aligned}\lambda \mathbf{x}\cdot\mathbf{x} &= A\mathbf{x}\cdot\mathbf{x} \overset{A \text{ Hermitian}}{=} \mathbf{x}\cdot A\mathbf{x} \\ &= \mathbf{x}\cdot\lambda\mathbf{x} \\ &\overset{\text{rule for complex inner product}}{=} \overline{\lambda}\mathbf{x}\cdot\mathbf{x}\end{aligned}$$

and so $\lambda = \overline{\lambda}$. This shows that all eigenvalues are real. Now let $\mathbf{x}, \mathbf{y}, \mu$ and λ be given as above.

$$\begin{aligned}\lambda(\mathbf{x}\cdot\mathbf{y}) = \lambda\mathbf{x}\cdot\mathbf{y} = A\mathbf{x}\cdot\mathbf{y} = \\ \mathbf{x}\cdot A\mathbf{y} = \mathbf{x}\cdot\mu\mathbf{y} \\ \overset{\text{rule for complex inner product}}{=} \overline{\mu}(\mathbf{x}\cdot\mathbf{y}) \\ = \mu(\mathbf{x}\cdot\mathbf{y})\end{aligned}$$

and so

Since $\lambda \neq \mu$, it follows $\mathbf{x}\cdot\mathbf{y} = 0$.

19. Certainly not. $\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}$

20. eigenvectors:

$$\left\{\begin{pmatrix} -\frac{1}{6}\sqrt{6} \\ -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{2}\sqrt{3} \end{pmatrix}\right\} \leftrightarrow 6,$$
$$\left\{\begin{pmatrix} -\frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \\ 0 \end{pmatrix}\right\} \leftrightarrow 12,$$
$$\left\{\begin{pmatrix} \frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \\ \frac{1}{3}\sqrt{3} \end{pmatrix}\right\} \leftrightarrow 18.$$

The matrix U has these as its columns.

34. eigenvectors:

$$\left\{\begin{pmatrix} -\frac{1}{3}\sqrt{3} \\ 0 \\ \frac{1}{3}\sqrt{2}\sqrt{3} \end{pmatrix}\right\} \leftrightarrow 0,$$
$$\left\{\begin{pmatrix} \frac{1}{3}\sqrt{3} \\ -\frac{1}{2}\sqrt{2} \\ \frac{1}{6}\sqrt{6} \end{pmatrix}\right\} \leftrightarrow 1,$$

$\left\{\begin{pmatrix} \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} \\ \frac{1}{6}\sqrt{6} \end{pmatrix}\right\} \leftrightarrow 2.$ The columns are these vectors.

37. If A is given by the formula, then

$$A^T = U^T D^T U = U^T DU = A$$

Next suppose $A = A^T$. Then by the theorems on symmetric matrices, there exists an orthogonal matrix U such that

$$UAU^T = D$$

for D diagonal. Hence

$$A = U^T DU$$

39. There exists U unitary such that $A = U^* TU$ such that T is uppser triangular. Thus A and T are similar. Hence they have the same determinant. Therefore, $\det(A) = \det(T)$, but $\det(T)$ equals the product of the entries on the main diagonal which are the eigenvalues of A.

40. $\begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$

41. $y = -0.125x^2 + 1.425x + 0.925$

46. Find an orthonormal basis for the spans of the following sets of vectors.

a) $(3, -4, 0), (7, -1, 0), (1, 7, 1)$.
$\begin{pmatrix} 3/5 \\ -4/5 \\ 0 \end{pmatrix}, \begin{pmatrix} 4/5 \\ 3/5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

b) $(3, 0, -4), (11, 0, 2), (1, 1, 7)$
$\begin{pmatrix} \frac{3}{5} \\ 0 \\ -\frac{4}{5} \end{pmatrix}, \begin{pmatrix} \frac{4}{5} \\ 0 \\ \frac{3}{5} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

c) $(3, 0, -4), (5, 0, 10), (-7, 1, 1)$
$\begin{pmatrix} 3/5 \\ -4/5 \\ 0 \end{pmatrix}, \begin{pmatrix} 4/5 \\ 3/5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

47. $\begin{pmatrix} \frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \end{pmatrix}, \begin{pmatrix} \frac{3}{10}\sqrt{2} \\ -\frac{2}{5}\sqrt{2} \\ \frac{1}{2}\sqrt{2} \end{pmatrix}, \begin{pmatrix} \frac{7}{15}\sqrt{3} \\ -\frac{1}{15}\sqrt{3} \\ -\frac{1}{3}\sqrt{3} \end{pmatrix}$

49. $\begin{pmatrix} \frac{1}{5}\sqrt{5} \\ 0 \\ \frac{2}{5}\sqrt{5} \end{pmatrix}, \begin{pmatrix} -\frac{3}{35}\sqrt{5}\sqrt{14} \\ \frac{1}{14}\sqrt{5}\sqrt{14} \\ \frac{3}{70}\sqrt{5}\sqrt{14} \end{pmatrix}$

51. It satises the properties of an inner product. Note that

$$\begin{aligned} \overline{\text{trace}\,(AB^*)} &= \overline{\sum_i \sum_k A_{ik} \overline{B_{ik}}} \\ &= \sum_k \sum_i \overline{A_{ik}} B_{ik} \\ &= \text{trace}\,(BA^*) \end{aligned}$$

So

$$\overline{(A,B)_F} = (B,A)_F$$

The product is obviously linear in the rst argument. If $(A,A)_F = 0$, then

$$\sum_i \sum_k A_{ik}\overline{A_{ik}} = \sum_{i,k} |A_{ik}|^2 = 0$$

52. From the singular value decomposition,

$$\begin{aligned} U^*AV &= \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}, \\ A &= U\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}V^* \end{aligned}$$

Then

$$\begin{aligned} & \operatorname{trace}(AA^*) \\ = & \operatorname{trace}\left(U\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}V^* \cdot \right. \\ & \left. V\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}U^*\right) \\ = & \operatorname{trace}\left(U\begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}U^*\right) \\ = & \operatorname{trace}\begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix} = \sum_j \sigma_j^2 \end{aligned}$$

53. $\operatorname{trace}(AB) = \sum_i \sum_k A_{ik}B_{ki}$,
$\operatorname{trace}(BA) = \sum_i \sum_k B_{ik}A_{ki}$. These give the same thing. Now

$$\begin{aligned} \operatorname{trace}(A) &= \operatorname{trace}\left(S^{-1}BS\right) \\ &= \operatorname{trace}\left(BSS^{-1}\right) = \operatorname{trace}(B). \end{aligned}$$

C.14 EXERCISES 471

1. $\begin{pmatrix} 0.39 \\ -0.09 \\ 0.53 \end{pmatrix}$

2. $\begin{pmatrix} 5.3191 \times 10^{-2} \\ 7.4468 \times 10^{-2} \\ 0.71277 \end{pmatrix}$

5. $\begin{pmatrix} 0.14394 \\ 0.93939 \\ 0.2803 \end{pmatrix}$

6. $\begin{pmatrix} 0.20521 \\ 0.11726 \\ -2.6059 \times 10^{-2} \end{pmatrix}$

7. It indicates that they are no good for doing it.

C.15 EXERCISES 501

1. The actual largest eigenvalue is 8 with corresponding eigenvector $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$

4. The largest eigenvalue is -16 and an eigenvector is $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$.

5. Eigenvalue near $-1.35 : \lambda = -1.341$,

$$\begin{pmatrix} 1.0 \\ -0.45606 \\ -0.47632 \end{pmatrix}$$

Eigenvalue near 1.5: $\lambda = 1.6790$,

$$\begin{pmatrix} 0.86741 \\ 5.5869 \\ -3.5282 \end{pmatrix}$$

Eigenvalue near 6.5: $\lambda = 6.662$,

$$\begin{pmatrix} 4.4052 \\ 3.2136 \\ 6.1717 \end{pmatrix}$$

8. Eigenvalue near $-1 : \lambda = -0.70369$,

$$\begin{pmatrix} 3.3749 \\ -1.2653 \\ 0.15575 \end{pmatrix}$$

Eigenvalue near $.25 : \lambda = 0.18911$,

$$\begin{pmatrix} -0.24220 \\ -0.52291 \\ 1.0 \end{pmatrix}$$

Eigenvalue near $7.5 : \lambda = 7.5146$,

$$\begin{pmatrix} 0.34692 \\ 1.0 \\ 0.60692 \end{pmatrix}$$

10.

$$\left|\lambda - \frac{22}{3}\right| \leq \frac{1}{3}\sqrt{2}$$

12. From the bottom line, a lower bound is -10 From the second line, an upper bound is 12.

ENDNOTES

[9] This word has 9 syllables. It is a little like the name of a volcano in Iceland.

[10] Recall that for a complex number, $x+iy$, the complex conjugate, denoted by $\overline{x+iy}$ is defined as $x-iy$.

INDEX

N

O

P

To see Chapter 16-18 download

Elementary Linear Algebra: Part III