Contents

1 Gilbert Strang Video Lectures from MIT Opencourse, Spring 2005

1.1 Geometry of Linear Equations

We can tackle the problem of finding a solution to a linear system in three different (but equivalent) ways. Consider the following example: Solve the equations

$$x + 2y = 3$$
$$2x - y = 7$$

for x and y. The first approach to solving this problem is to think of the xy-plane where the solution is the point (x, y) corresponding to the intersection of the two lines given by the above equations. The second, and more practical way, is to simply write the two equations in matrix notation and solve it. The third, and less obvious way would be to think of the pair (x, y) for which the linear combination

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

works (picture this geometrically in your mind and compare this to the line intersection picture earlier). This third perspective offers some intuitive insight into concepts that will arise later, such as how when $x, y \in \mathbb{R}$ then the above linear combination spans the whole xy-plane, but only a unique combination gives us the point we want.

The next obvious question to ask is whether there's a solution to $\mathbf{Ax} = \mathbf{b}$ for any b? In other words, do the linear combination of the columns fill 2-dimensional space? For the case shown above, the answer is yes, because the matrix $\mathbf{A}$ is a non-singular (i.e. invertible) matrix. This obviously extend to the n-dimensional case. We can thus think of the multiplication $\mathbf{Ax}$ in two ways: (1) the dot product of each row of $\mathbf{A}$ with $\mathbf{x}$, or (2) the linear combination consisting of the sum over i of the ith column of $\mathbf{A}$ with the ith element of $\mathbf{x}$. **So Ax is a combination of the columns of A.**

1.2 Strang Video Lecture: Elimination with Matrices

Elimination is the process through which we take the augmented matrix $\mathbf{A}|\mathbf{b}$ from the equation $\mathbf{Ax} = \mathbf{b}$ and turn $\mathbf{A}$ into an upper triangular matrix by subtracting from a row the right multiple of the row above it, starting with the second row. The diagonal elements of the resulting upper triangular matrix (which cannot be zero) are called **pivots** and their product gives the determinant of the matrix. It is the most common computation in scientific computing. Once we have obtained the upper-triangular matrix, we start with the last row and substitute the obtained value into the row above it until we have all the elements of $\mathbf{x}$. This process is called **back substitution**. Here's an example with two steps:

$$\begin{pmatrix} 1 & 2 & 1 & | & 2 \\ 3 & 8 & 1 & | & 12 \\ 0 & 4 & 1 & | & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & | & 2 \\ 0 & 2 & -2 & | & 6 \\ 0 & 0 & 5 & | & -10 \end{pmatrix}$$

Remember that elimination is a **row operation**. Earlier we learned that **a matrix times a column gave us a combination of the columns of the matrix**. Similarly, **a row times a matrix gives us a combination of the rows of the matrix**. This means that for example the operation of subtracting 3 times the first row from the second row of a 3×3 matrix $\mathbf{A}$ (leaving the first and third row unchanged) is equivalent to the following matrix multiplication:

$$\begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{A} = \begin{pmatrix} A_1 \\ A_2 - 3A_1 \\ A_3 \end{pmatrix}$$

A second matrix can now be appended to the beginning of the above multiplication accordingly, giving us $\mathbf{E}_{32}(\mathbf{E}_{21}\mathbf{A}) = (\mathbf{E}_{32}\mathbf{E}_{21})\mathbf{A}$ where $\mathbf{E}_{32}\mathbf{E}_{21}$ is our **elimination matrix** that makes first a_{21} and then a_{32} zero (a_{31} was already zero, otherwise we would also need $\mathbf{E}_{31}$) as well. (Remember that matrix multiplication is associative but not commutative.)

A matrix $\mathbf{P}$ that switches the rows of $\mathbf{A}$ to give you $\mathbf{PA}$ is called a **permutation matrix**. By switching this to $\mathbf{AP}$ we would change the columns of $\mathbf{A}$ instead of its rows. For a 2×2 matrix we have $\mathbf{P} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right)$.

1.3 Multiplication and Inverses

Matrix multiplication can be thought of in different ways. For the equation $\mathbf{AB} = \mathbf{C}$, we can think of $\mathbf{C} = (c_{ij})$ where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ (or simply $\mathbf{A}_i \cdot \mathbf{B}^j$), where $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$. You can also think of $\mathbf{AB}^j = \mathbf{C}^j$ for $j = 1, 2, \ldots, p$. In other words, each column of $\mathbf{C}$ is a linear combination of the columns of $\mathbf{A}$. Alternatively, you can think of each row of $\mathbf{C}$ as a combination of the rows of $\mathbf{B}$, i.e. $\mathbf{C}_i = \mathbf{A}_i\mathbf{B}$. Finally, you can think of $\mathbf{AB}$ as the sum of columns of $\mathbf{A}$ times rows of $\mathbf{B}$. Here's an example:

$$\begin{pmatrix} 2 & 7 \\ 3 & 8 \\ 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 6 \end{pmatrix} + \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix}$$

Note how this approach shows the matrix as the sum of two matrices with the same size as matrix $\mathbf{C}$.

It is also possible to divide the matrices $\mathbf{A}$ and $\mathbf{B}$ into blocks and do **block multiplication**.

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_3 & \mathbf{B}_4 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{pmatrix}$$

where for example $\mathbf{C}_1 = \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_3$.

Let's now think of the problem of inverses of square matrices. The main question is how do I identify invertible (non-singular) matrices, i.e. $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{AA}^{-1}$. We already know that when $\det(A) = 0$ it follows that $\mathbf{A}$ is singular. But we can also think of it as follows: **A square matrix $\mathbf{A}$ has no inverse if I can find a nonzero vector x such that $\mathbf{Ax} = 0$.** In effect what I'm saying here is that when $\mathbf{Ax} = 0$ then the columns of $\mathbf{A}$ are linearly dependent.

Gauss-Jordan is a method for solving multiple equations at once. Think of this in terms of finding $\mathbf{A}^{-1}$, i.e. we want a $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{I}$ or $\mathbf{AB}^j = \mathbf{I}^j = \mathbf{e}_j$ for all $j = 1, 2, \ldots, n$. Take the simple example of

$$\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which really consists of two simultaneous equations given by $\mathbf{A}\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) = \mathbf{e}_1$ and $\mathbf{A}\left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) = \mathbf{e}_2$. We can solve each one separately by augmenting the right-hand sides to $\mathbf{A}$ and using elimination, or we can augment both simultaneously. We begin by making the matrix upper-triangular, just as in elimination, but we then make the matrix diagonal by adding to upper rows multiples of lower rows.

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

In short, we have $\mathbf{E}(\mathbf{A}|\mathbf{I}) = (\mathbf{I}|\mathbf{A}^{-1})$.

1.4 Factorization into $\mathbf{A} = \mathbf{LU}$

From the equation $\mathbf{AA}^{-1} = (\mathbf{AA}^{-1})^T = (\mathbf{A}^{-1})^T\mathbf{A}^T = \mathbf{I}$ it follows that $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)i$. Assume that $\mathbf{A}$ is some 3×3 matrix and that through the process of elimination we get $\mathbf{E}_{32}\mathbf{E}_{31}\mathbf{E}_{21}\mathbf{A} = \mathbf{U}$ (assuming you don't have to do any row exchanges as part of the elimination process). We want to find $\mathbf{L}$ such that $\mathbf{A} = \mathbf{LU}$ (where $\mathbf{L}$ is a lower-triangular matrix and $\mathbf{U}$ is an upper-triangular matrix. It follows that $\mathbf{L} = \mathbf{E}_{21}^{-1}\mathbf{E}_{31}^{-1}\mathbf{E}_{32}^{-1}$, where finding these inverses is really easy, for example let

$$\mathbf{E} = \mathbf{E}_{32}\mathbf{I}\mathbf{E}_{21} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 10 & -5 & 1 \end{pmatrix}$$

and the we can obtain $\mathbf{L}$ as follows:

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix}$$

So for $\mathbf{A} = \mathbf{LU}$, if there are no row exchanges, the multipliers go directly into $\mathbf{L}$. So finding $\mathbf{L}$ is just that easy. This property of $\mathbf{L}$ greatly enhances computational ease for large matrices.

1.5   Transposes, Permutations

When row exchanges must be made we must start by using permutation matrices before we can proceed with elimination. Now the equation $\mathbf{A} = \mathbf{LU}$ becomes $\mathbf{PA} = \mathbf{LU}$ for any invertible $\mathbf{A}$. Think of the permutation $\mathbf{P}$ as the identity matrix with reordered rows. A nice thing about permutation matrices is that $\mathbf{P}^{-1} = \mathbf{P}^T$.

Recall that a **symmetric matrix** is one that is equal to its transpose. This is particularly helpful when one considers that for any matrix $\mathbf{R}$, the matrix $\mathbf{R}^T\mathbf{R}$ is always symmetric. This is true because $(\mathbf{R}^T\mathbf{R})^T = \mathbf{R}^T\mathbf{R}$.

1.6   Column Space and Nullspace

We know that $\mathbb{R}^2$ is a **vector space**. The subspaces of $\mathbb{R}^2$ are itself, all the lines through the origin, and the origin itself. By taking all the linear combinations of the columns of a matrix, we get what's called a **column space**. We can now go back to the equation $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is not necessarily square anymore and say that this equation can be solved for any $\mathbf{b}$ that is in the column space of $\mathbf{A}$. On the other hand, the **nullspace** of $\mathbf{A}$, $N(\mathbf{A})$, is the space spanned by the vectors $\mathbf{x}$ for which $\mathbf{Ax} = 0$. For an $m \times n$ matrix, the column space is in $\mathbb{R}^m$ while the nullspace is in $\mathbb{R}^n$. It turns out that for all the vectors $x \in N(\mathbf{A})$ we have $\mathbf{Ax} = 0$. So we can also define the nullspace as the collection of all vectors $\mathbf{x}$ for which $\mathbf{Ax} = 0$. It is easy to show that the nullspace is indeed a subspace. The obvious extension is now to ask whether the collection of all $\mathbf{x}$ for which $\mathbf{Ax}$ equals some non-zero $\mathbf{b}$ is also a subspace. The simple answer is no, because the zero vector is not one of them and a subspace must contain the zero vector.

1.7   Solving Ax = 0: Pivot Variables

We can apply what we learned about elimination to any rectangular matrix, and write the matrix in what's called the **echelon form**. Take for example the equation $\mathbf{Ax} = 0$ where

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{U}$$

Notice that $\mathbf{A}$ has now two pivots. We call the number of pivots of the matrix $\mathbf{A}$ written in echelon form the **rank** of $\mathbf{A}$. The solution $\mathbf{x}$ to the above equation can be found by making some arbitrary choices for the free variables and determining the value of the pivot variables. But notice that $\mathbf{A}^2 = 2\mathbf{A}^1$. This tells you that $(-2, 1, 0, 0)^T \in N(\mathbf{A})$ (so every multiple of it is also a solution). Since the free variables here had the values 0 and 1, then let them be 1 and 0 now and you get $(-2, 0, 2, 1)^T \in N(\mathbf{A})$. So the nullspace of $\mathbf{A}$ is any linear combination of the two vectors given above.

Similar to the echelon form, we also have the **reduced echelon form** which has zeros above as well as below the pivots. For matrix $\mathbf{A}$ we have (after making the pivots equal 1)

$$\mathbf{R} = \begin{pmatrix} \boxed{1} & 2 & 0 & -2 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

So now we can find the solutions to $\mathbf{Rx} = 0$. But notice that the solutions we found from $\mathbf{Ux} = 0$ are already apparent by looking at the matrix $\mathbf{R}$. By switching the order of the columns in $\mathbf{R}$ so that the pivot columns come first followed by the free columns we can represent $\mathbf{R}$ with the general form

$$\mathbf{R} = \begin{pmatrix} \mathbf{I} & \mathbf{F} \\ 0 & 0 \end{pmatrix}$$

which shows that $\mathbf{R}$ has rank$(\mathbf{A}) = r$ pivot columns, r pivot rows, and $n - r$ free columns. We want to create a **nullspace matrix**, i.e. a matrix $\mathbf{N}$ such that $\mathbf{RN} = 0$. It turns out that $\mathbf{N}$ is just the matrix $(-\mathbf{F}\ \mathbf{I})^T$ (note that the two identity matrices in $\mathbf{R} = (\mathbf{I}\ \mathbf{F})$ and $\mathbf{N} = (-\mathbf{F}\ \mathbf{I})^T$ are not necessarily of the same size; $\mathbf{I}$ in the first equation is has a size equal to the number of rows of $\mathbf{F}$ and $\mathbf{I}$ in the second equation has a size equal to the number of columns of $\mathbf{F}$. Notice that we can write $\mathbf{x}_{\text{pivot}} = -\mathbf{F}\mathbf{x}_{\text{free}}$ because

$$\mathbf{Rx} = \begin{pmatrix} \mathbf{I} & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{\text{pivot}} \\ \mathbf{x}_{\text{free}} \end{pmatrix} = 0$$

We leave it up to you to demonstrate that if the matrix $\mathbf{A}$ was the inverse of the matrix given above, we would have

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 6 & 8 \\ 2 & 8 & 10 \end{pmatrix} \Rightarrow \mathbf{U} = \begin{pmatrix} \boxed{1} & 2 & 3 \\ 0 & \boxed{2} & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \mathbf{R} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and we can see that $\text{rank}(\mathbf{A}) = 2$ (which leads us to the conjecture $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$). Furthermore, from noticing that $\mathbf{A}^3 = \mathbf{A}^1 + \mathbf{A}^2$ we can quickly expect $\mathbf{N}(A) = c(-1, -1, 1)^T$ which from looking at $\mathbf{R}$ is nothing but $c(-\mathbf{F}\ \mathbf{I})^T$ for some $c \in \mathbb{R}$.

1.8   Solving Ax = b: Row Reduced Form R

We are finally at the point of finding a sweeping solution to the linear equation $\mathbf{Ax} = \mathbf{b}$. We do elimination on the augmented matrix $(\mathbf{Ab})$ to get

$$\begin{pmatrix} 1 & 2 & 2 & 2 & | & b_1 \\ 2 & 4 & 6 & 8 & | & b_2 \\ 3 & 6 & 8 & 10 & | & b_3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 2 & 2 & | & b_1 \\ 0 & 0 & 2 & 4 & | & b_2 - 2b_3 \\ 0 & 0 & 0 & 0 & | & b_3 - b_2 - b_1 \end{pmatrix}$$

where the last equation gives $b_3 - b_2 - b_1 = 0$, which is the main condition for solvability. This condition goes hand in hand with the fact that $\mathbf{A}_3 - \mathbf{A}_2 - \mathbf{A}_1 = 0$. Earlier we learned that for $\mathbf{Ax} = \mathbf{b}$ to be solvable, then $\mathbf{b}$ must be in the column space of $\mathbf{A}$, $C(\mathbf{A})$. Now we know that if a combination of the rows of $\mathbf{A}$ gives the zero row, then the same combination of $\mathbf{b}$ must give zero.

In our quest for a complete solution to $\mathbf{Ax} = \mathbf{b}$ we can first find a particular solution by setting all free variables equal to zero and solving for the pivot variables. For the above case we have $x_2 = x_4 = 0$ leaving us with $x_1 + 2x_3 = 1$ and $2x_3 = 3$ which gives $x_1 = -2$ and $x_3 = 3/2$. The complete solution will then consist of this particular solution, plus any solution out of the nullspace, in short $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$. This is because summing $\mathbf{Ax}_p = \mathbf{b}$ and $\mathbf{Ax}_n = 0$ gives $\mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) = \mathbf{b}$. For the above case we have

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n = \begin{pmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

For an $m \times n$ matrix $\mathbf{A}$ of rank r so far we now that $r \le m$ and $r \le n$. When $r = n < m$ we say $\mathbf{A}$ is **full column rank**. In this case we have no free variables, i.e. $N(\mathbf{A}) = 0$ and hence the solution to $\mathbf{Ax} = \mathbf{b}$ is the unique solution $\mathbf{x}_p$ (if there is one, i.e. if $\mathbf{b} \in C(\mathbf{A})$). In other words we have either zero or one solution. When $r = m < n$ we call $\mathbf{A}$ **full row rank**, i.e. we have m pivots and we can solve $\mathbf{Ax} = \mathbf{b}$ for every $\mathbf{b}$, because we are left with $n - m = n - r$ free variables. In other words, we have infinitely many solutions. When $r < m$ and $r < n$, we have either no solutions or infinitely many of them. When $r = n = m$ we have a square matrix that's **full rank**, is invertible and whose reduced echelon form is just $\mathbf{I}$. In this case, since the $r = m$ then we can solve for every $\mathbf{b}$ and since $r = n$ then there's a unique solution.

1.9   Independence, Basis, and Dimension

Any set of vectors where at least one is the zero vector is linearly dependent. As we learned before, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are the columns of $\mathbf{A}$ then they are independent when $N(\mathbf{A}) = 0$, i.e. $\text{rank}(\mathbf{A}) = n$. The rank of a matrix $\mathbf{A}$ was previously defined as the number of pivot columns in $\mathbf{A}$; it can now be described as the **dimension** of the column space of $\mathbf{A}$, $C(\mathbf{A})$, which is the number of vectors required to span $C(\mathbf{A})$, also called **bases**. Moreover, the dimensions of the nullspace of $\mathbf{A}$ is the number of free variables.

1.10   The Four Fundamental Subspaces

The four fundamental subspaces are (1) the column space, $C(\mathbf{A})$, (2) the nullspace, $N(\mathbf{A})$, (3) the **row space** (all the combinations of the columns of $\mathbf{A}^T$, $C(\mathbf{A}^T)$, and (4) the nullspace of $\mathbf{A}^T$, $N(\mathbf{A}^T)$ also called the **left nullspace** of $\mathbf{A}$.

It should be obvious that for an $m \times n$ matrix $\mathbf{A}$, $C(\mathbf{A}) \in \mathbb{R}^m$, $N(\mathbf{A}) \in \mathbb{R}^n$ and vice versa for $\mathbf{A}^T$. Of interest is to find a systematic way of constructing a basis for each of those spaces. Starting with $C(\mathbf{A})$, we could simply use the pivot columns of $\mathbf{A}$ as a basis for it and conclude $\dim(C(\mathbf{A})) = \text{rank}(\mathbf{A}) = r$. It turns out that $\dim(C(\mathbf{A}^T)) = \text{rank}(\mathbf{A}) = r$ as well. For $N(\mathbf{A})$ we can use the special solutions we obtained by choosing the free variables as the basis, and we have $n - r$ of them, so $\dim(N(\mathbf{A})) = n - r$. We will expand on these basic results, but let's start by showing you an example of how row operation does not preserve the column space:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix} \rightarrow \mathbf{R} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{F} \\ 0 & 0 \end{pmatrix}$$

Note that in the above example $C(\mathbf{A}) \neq C(\mathbf{R})$, but we argue that $\mathbf{A}$ and $\mathbf{R}$ have the same row space, because $\mathbf{R}$ is obtained by performing row operations on $\mathbf{A}$ and so we stay in the basis of $\mathbf{A}$. Moreover, a basis for the row space of $\mathbf{R}$ (and hence $\mathbf{A}$) is the first two rows of $\mathbf{R}$.

Now let's study $N(\mathbf{A}^T)$, i.e. all vectors $\mathbf{y}$ for which $\mathbf{A}^T \mathbf{y} = 0$, which we call the left nullspace because we can rewrite $\mathbf{y}^T \mathbf{A} = 0$. Remember how we used Gauss-Jordan to find an inverse of a square matrix by augmenting the identity matrix to it and finding the reduced row echelon form. We will do a similar thing here, except $\mathbf{A}$ is $m \times n$, and we get $(\mathbf{A} \ \mathbf{I}) \rightarrow (\mathbf{R} \ \mathbf{E})$ where $\mathbf{E}$ is a matrix that contains a record of what we did to get $\mathbf{R}$, i.e. $\mathbf{EA} = \mathbf{R}$. To find $\mathbf{E}$ for the previous example, just start with the identity matrix and make the changes that represent the row operations:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \mathbf{E}$$

We can use the matrix $\mathbf{E}$ to find the basis for the left nullspace of $\mathbf{A}$ and show that $\dim(N(\mathbf{A}^T)) = m - r$. Note that the last row of $\mathbf{E}$ corresponds to the combination of the rows of $\mathbf{A}$ that gives me the zero vector, so $\mathbf{E}_3$ is the basis for the left nullspace. **So the row space and the nullspace are in $\mathbb{R}^n$ and their dimensions add up to n, and the column space and the left nullspace are in $\mathbb{R}^m$ and their dimensions add up to m.**

1.11   Matrix Spaces; Rank 1; Small World Graphs

We can extend what we learned about vector spaces by thinking of a vector space as being all 3×3 matrices. Note that we can add them and multiply them by constants, so they are a vector space (the fact that we can multiply them too is irrelevant for now). So now examples of subspaces are upper triangular matrices, symmetric matrices, or their intersection, i.e. diagonal matrices. We can also talk about bases for these subspaces. For example, diagonal matrices have three bases and hence have a dimension of 3. So we want to know what the bases are for these subspaces. Let's first find a basis for M, the vector space consisting of all 3×3 matrices. A basis for M could look like

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

of which we have 9, i.e. $\dim(M) = 9$. For symmetric matrices S, the dimension is 6, so is the dimension of upper-triangular matrices, U. For diagonal matrices, $S \cap U$, we have a dimension of 3. Note that $S \cup U$ is not a subspace to begin with. On the other hand $S + U = \{s + u | s \in S \text{ and } u \in U\}$ gives me M again, and hence $\dim(S + U) = 9$. So we note that $\dim(S) + \dim(U) = \dim(S + U) + \dim(S \cap U)$.

Let's look at another example. Consider the differential equation $d^2y/dx^2 + y = 0$. The two solutions that we can think of are $\cos x$ and $\sin x$ and all the other solutions are linear combinations of the two. So think of the nullspace as being the solution to the above equation, and $\cos x$ and $\sin x$ as the bases for the nullspace. Let's come back to matrices and think again about the rank of a matrix. Consider the matrix $\mathbf{A}$ given by

$$\begin{pmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 4 & 5 \end{pmatrix}$$
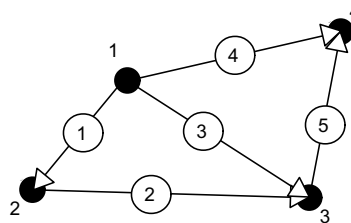
where it should be obvious that $\dim(C(\mathbf{A})) = 1$ which as we learned must also equal $\dim(C(\mathbf{A}^T)) = \text{rank}(\mathbf{A})$. But is there a more obvious way to see that? The answer lies in writing $\mathbf{A}$ as the multiplication shown on the right-hand side above, which leads us to say that every matrix of rank 1 can be thought of as some column times some row, i.e. $\mathbf{A} = \mathbf{uv}^T$ where $\text{rank}(\mathbf{A}) = 1$. We can therefore think of rank 1 matrices as the building block for every matrix, e.g. a rank 4 matrix can be written as the sum of 4 rank 1 matrices. Note that the set consisting of all matrices of rank 2 for example is not a subspace. Consider another example now: let's say we're interested in the subspace consisting of all vectors $\mathbf{v} \in \mathbb{R}^4$ such that the elements of $\mathbf{v}$ sum to 0 (check that it is in fact a subspace). We want to know the dimension of this subspace. One way to think of it is that since the four combinations are not linearly independent, then the dimension must be 3. To confirm this think of a matrix $\mathbf{A}$ whose nullspace is the above subspace. Then $\mathbf{A}$ must be the matrix with 4 columns whose elements are all 1. Since $\text{rank}(\mathbf{A}) = 1$ then we have $\text{rank}(N(\mathbf{A})) = 4 - 1 = 3$. We could find a basis for $N(\mathbf{A})$ by setting the values for the free variables and then finding the pivot variable. A basis could be

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

If $\mathbf{A}$ consists of only one column, then $C(\mathbf{A}) = \mathbb{R}$ and $N(\mathbf{A}^T) = 0$. Note how $n = 4 = 3 + 1$ and $m = 1 = 1 + 0$ and everything works out.

1.12 Graphs, Networks, Incidence Matrices

Consider a graph whose nodes are connected by directed edges which we have numbered. We can represent this graph with a matrix, called an **incidence matrix** by letting a given row of the matrix represent an edge and every column a node. The entry of the matrix consist of 1, 0, and -1 only with a_{ij} equals 1 when edge i goes into node j, 0 when edge i does not connect to node j, and -1 when edge i leaves node j. Take the following graph and corresponding matrix:

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

If we stop at the third row and take a look at the matrix we would notice that the rows are not independent anymore, which on the graph corresponds to having a **loop**. This is an example of a **sparse** matrix, because it has a large number of zeros (which gets larger the larger the matrix gets). Let's first see if the columns of $\mathbf{A}$ are independent, i.e. if the nullspace is just the zero vector, by studying the equation $\mathbf{Ax} = 0$. We call $\mathbf{x} = (x_1, x_2, x_3, x_4)$ the **potentials at the nodes**, and hence $\mathbf{Ax}$ (given the structure of $\mathbf{A}$) gives me the **potential differences** across the edges. The nullspace is where the potential differences are all zero. It's easy to see that the nullspace consists of $c(1, 1, 1, 1)^T$ and hence $\text{rank}(\mathbf{A}) = 3$. It follows that $\dim(\mathbf{A}^T) = 5 - 3 = 2$ and for $\mathbf{A}^T\mathbf{y} = 0$ we call $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)$ the **currents on edges** (where the currents are the directions on the edges). The relation between current and potential difference is called **Ohm's law**, which says that the current on an edge is some number times the potential drop. It's a change in potential that makes a current happen and it's Ohm's law that says how much current happens. Looking at the rows of $\mathbf{A}^T\mathbf{y}$ you'll see that each row represents total flow on its corresponding node. So instead of using row reduction, we can look at the graph to find the nullspace of $\mathbf{A}^T$. Two bases for that nullspace would be $(1, 1, -1, 0, 0)$ and $(0, 0, 1, -1, 1)$, which are just currents around some loop (any loop works). You find them by choosing a loop, assigning a direction to it (clockwise

or counter-clockwise), and assigning the value of 1 to y_i when it flows in that direction, -1 when it flows in the opposite direction, and 0 when it isn't part of the loop. Note how the sum of the two bases above gives us $(1, 1, 0, -1, 1)$, which is the current around the big loop. Another interesting observation is that we can use the first basis to find the pivot columns of $\mathbf{A}^T$, i.e. the independent rows of $\mathbf{A}$. I expect 3 pivot columns, since $\text{rank}(\mathbf{A}) = 3$, so I can pick the first two columns and the forth one. On the graph, this selection corresponds to a subgraph without a loop, also known as a **tree**. Finally, from $\dim(N(\mathbf{A}^T)) = m - r$ you'll notice that $\dim(N(\mathbf{A}^T))$ is the number of loops, $m = 5$ is the number of edges, and $r = n - 1$ where $n = 4$ is the number of nodes. This gives us a count that applies to every graph

$$\# \text{ nodes} - \# \text{ edges} + \# \text{ loops} = 1$$

called **Euler's formula**. We used linear algebra to prove Euler's formula, which otherwise would have required a topological proof.
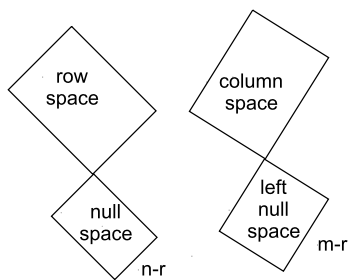
1.13 Quiz 1 Review

So the row space and the nullspace are in $\mathbb{R}^n$ and their dimensions add up to n, and the column space and the left nullspace are in $\mathbb{R}^m$ and their dimensions add up to m. An important fact to remember is that $N(\mathbf{A}) = 0$ if and only if $N(\mathbf{A}^T) = 0$.

Consider an example now: Let $\mathbf{B} = \mathbf{CA}$ where $\mathbf{C}$ is an invertible square matrix that's $m \times m$, and $\mathbf{B}$ and $\mathbf{A}$ are $m \times n$. We're interested in $N(\mathbf{B}) \in \mathbb{R}^n$. Notice that since $\mathbf{C}$ is invertible, then $N(\mathbf{CA}) = N(\mathbf{A})$, i.e. **multiplying by an invertible matrix can't change the nullspace**.

1.14 Orthogonal Vectors and Subspaces

The illustration below is a summary of what we learned so far, as well as an insight into a new topic, namely how the row space and nullspace are orthogonal to each other, as are the column space and left nullspace. Let's first see why it is that

$\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $\mathbf{x}^T\mathbf{y} = 0$. We can show this using the fact that $||\mathbf{x}||^2 + ||\mathbf{y}||^2 = ||\mathbf{x}+\mathbf{y}||^2$ and rewriting it as $\mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} = (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y})$ and expanding on the right side of the equation to get $0 = \mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{x} = 2\mathbf{x}^T\mathbf{y}$. So the zero vector is orthogonal to any vector. For subspaces, orthogonality means that every vector in one subspace must be orthogonal to every vector in the other subspace. From the equation $\mathbf{Ax} = 0$, it is easy to see why the row space of $\mathbf{A}$ is orthogonal to $\mathbf{x} \in N(\mathbf{A})$. The properties of orthogonality combined with the fact that the dimensions of the row space and the nullspace always add up to n make these two spaces **orthogonal complements**. This amazing result can be considered part 1 of the fundamental theorem of linear algebra.

Consider now the problem of "solving" the equation $\mathbf{Ax} = \mathbf{b}$ for when $m > n$ and there's no solution, e.g. $\mathbf{b}$ has measurement "error" or residuals. So by "solving" it we mean finding the best solution. It is obvious that elimination will fail us. We know the matrix $\mathbf{A}^T\mathbf{A}$ is a square and symmetric matrix. So let's have $\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$. We changed $\mathbf{x}$ to $\hat{\mathbf{x}}$ because we think of $\mathbf{x}$ as the solution to the original equation if it existed (but probably didn't) and I'm hoping that $\hat{\mathbf{x}}$ does exist. We can't say that $\mathbf{A}^T\mathbf{A}$ is always invertible. Here's an example,

$$\begin{pmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 9 \\ 3 & 27 \end{pmatrix}$$

and since $\text{rank}(\mathbf{A}) = 1$, we cannot expect $\text{rank}(\mathbf{A}^T\mathbf{A})$ to be any higher than 1. This leads to the important result that $N(\mathbf{A}^T\mathbf{A}) = N(\mathbf{A})$ and $\text{rank}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A})$. So $\mathbf{A}^T\mathbf{A}$ is invertible exactly if $\mathbf{A}$ has independent columns.

1.15   Projections onto Subspaces

When you project a vector $\mathbf{b}$ on a vector $\mathbf{a}$ you can think of the difference between $\mathbf{b}$ and its projection $\mathbf{p}$ as an "error" $\mathbf{e} = \mathbf{b} - \mathbf{p}$. We can let $\mathbf{p} = x\mathbf{a}$ and write the key equation $\mathbf{a}^T(\mathbf{b} - x\mathbf{a}) = 0$ which gives $x\mathbf{a}^T\mathbf{a} = \mathbf{a}^T\mathbf{b}$. So we have $\mathbf{p} = \mathbf{a}\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}$. Note that if you multiply $\mathbf{b}$ by some number then its new projection is also the old projection times the same number. So the projection is really just some matrix $\mathbf{P}$ such that $\operatorname{proj} \mathbf{b} = \mathbf{Pb}$. It's easy to see that $\mathbf{P} = \frac{\mathbf{aa}^T}{\mathbf{a}^T\mathbf{a}}$ where the denominator is a scalar, but the numerator is a matrix since it is obtained by multiplying a column by a row. Note how the column space of $\mathbf{P}$ is the line through $\mathbf{a}$ and the rank of $\mathbf{P}$ is 1. Furthermore, $\mathbf{P}$ is symmetric, which is a key property, and when we do the projection twice we stay put the second time, i.e. $\mathbf{P}^2 = \mathbf{P}$ ($\mathbf{P}$ is **idempotent**). Let's now step back a little and ask why we should project in the first place. The reason has to do once before with not being able to solve the equation $\mathbf{Ax} = \mathbf{b}$ and hence looking for the best solution. Think

Imagine now trying to project $\mathbf{b}$ on a plane spanned by the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, i.e. the column space of the matrix $\mathbf{A} = (\mathbf{a}_1 \ \mathbf{a}_2)$. Since $\mathbf{p}$ is on the plane, then I can write it as $\mathbf{p} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{A\hat{x}}$. So we want $\hat{x}$ such that $\mathbf{p} = \mathbf{A\hat{x}}$. I also know that $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - \mathbf{A\hat{x}}$ is perpendicular to the plane, and hence perpendicular to both $\mathbf{a}_1$ and $\mathbf{a}_2$. So I can say that $\mathbf{a}_1^T(\mathbf{b} - \mathbf{A\hat{x}}) = 0$ and $\mathbf{a}_2^T(\mathbf{b} - \mathbf{A\hat{x}}) = 0$, which can be combined into the matrix multiplication $\mathbf{A}^T(\mathbf{b} - \mathbf{A\hat{x}}) = \mathbf{A}^T\mathbf{e} = 0$. So $\mathbf{e}$ is in the nullspace of $\mathbf{A}^T$, i.e. $\mathbf{e}$ is perpendicular to $C(\mathbf{A})$. Let's now rewrite the equation as $\mathbf{A}^T\mathbf{A\hat{x}} = \mathbf{A}^T\mathbf{b}$, which gives $\hat{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ and hence $\mathbf{p} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{Pb}$. Notice the similarity between $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ and $\frac{\mathbf{aa}^T}{\mathbf{a}^T\mathbf{a}}$. Note that because $\mathbf{A}$ is not a square and invertible matrix, we can't just write $(\mathbf{A}^T\mathbf{A})^{-1} = \mathbf{A}^{-1}(\mathbf{A}^T)^{-1}$, and I can see how doing that would mess things up. But if $\mathbf{A}$ were a square and invertible matrix, then $\mathbf{b}$ would be in the column space of $\mathbf{A}$ and so its projection would be itself, i.e. $\mathbf{P} = \mathbf{I}$. So it still works. It's easy to see that $\mathbf{P}^T = \mathbf{P}$ (remember that the inverse of a symmetric matrix is also symmetric). It's also pretty easy to check that $\mathbf{P}^2 = \mathbf{P}$.

1.16   Projection Matrices and Least Squares

So far, we were able to turn the problem of not being able to solve $\mathbf{Ax} = \mathbf{b}$ into one of finding its best solution by using the equation $\mathbf{A}^T\mathbf{A\hat{x}} = \mathbf{A}^T\mathbf{b}$ called the **normal equations**. In the extreme cases, when $\mathbf{b}$ is in the column space of $\mathbf{A}$ then we get $\mathbf{PbA}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Ax} = \mathbf{Ax} = \mathbf{b}$, and when $\mathbf{b}$ is perpendicular to the column space of $\mathbf{A}$, i.e. $\mathbf{b}$ is in the nullspace of $\mathbf{A}^T$, then we get $\mathbf{Pb} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = 0$ because the last two terms give zero. We can thus decompose $\mathbf{b}$ into its orthogonal projections $\mathbf{p} = \mathbf{Pb}$ and $\mathbf{e} = (\mathbf{I} - \mathbf{P})\mathbf{b}$ and get $\mathbf{b} = \mathbf{p} + \mathbf{e}$. You can check to see that $\mathbf{e}$ is not only perpendicular to $\mathbf{p}$, but perpendicular to all of the vectors in the column space of $\mathbf{A}$. Alternatively, we can look at least squares as essentially the problem of minimizing the $||\mathbf{Ax} - \mathbf{b}||^2 = ||\mathbf{e}||^2$. At this point we could use calculus to solve this minimization problem, by taking the partial derivatives and setting them equal to zero. But the answer we would get is just the answer that linear algebra gave us, namely $\mathbf{A}^T\mathbf{A\hat{x}} = \mathbf{A}^T\mathbf{b}$.

Remember that we said that if $\mathbf{A}$ had independent columns, then $\mathbf{A}^T\mathbf{A}$ is invertible. This is the essential assumption that allowed us to find the least squares solution using linear algebra. We want to know what to do if $\mathbf{A}^T\mathbf{A}$ is not invertible. Suppose $\mathbf{A}^T\mathbf{Ax} = 0$, then we want to show that $\mathbf{x}$ must be zero, i.e. the nullspace of $\mathbf{A}^T\mathbf{A}$ is 0 which means it's invertible. An idea would be to consider $\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = 0 = (\mathbf{Ax})^T\mathbf{Ax}$, which means $||\mathbf{Ax}|| = 0$ and hence $\mathbf{Ax} = 0$. But since $\mathbf{A}$ has independent columns, we conclude $\mathbf{x} = 0$.

1.17   Orthogonal Matrices and Gram-Schmidt

The vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are **orthonormal** if $\mathbf{q}_i\mathbf{q}_j = 1$ when $i = j$ and $\mathbf{q}_i\mathbf{q}_j = 0$ otherwise. A matrix $\mathbf{Q}$ whose columns are orthonormal has the property that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, which is just an extension of the definition of orthonormality given above. When $\mathbf{Q}$ is a square matrix, we conventionally call it an **orthogonal matrix**, and we have $\mathbf{Q}^T = \mathbf{Q}^{-1}$ (this was also a property of permutation matrices and permutation matrices are a type of orthogonal matrices). Two good examples of $\mathbf{Q}$ is given by

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \ \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \text{ and } \frac{1}{3}\begin{pmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{pmatrix}.$$

Orthogonal matrices consisting of 1s and −1s such as the one above are of special interest. But to this day we don't have a way of knowing if for a given size n there exists such a matrix or not. Suppose we wanted to project onto $\mathbf{Q}$'s column

space, then the projection matrix is given by $\mathbf{P} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T = \mathbf{Q}\mathbf{Q}^T$. So our calculations are greatly simplified when using orthogonal matrices. Note that $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ if $\mathbf{Q}$ is square, because then I would be projecting onto the whole space. Also note that $(\mathbf{Q}\mathbf{Q}^T)(\mathbf{Q}\mathbf{Q}^T) = \mathbf{Q}\mathbf{Q}^T$. Our normal equations are now given by $\mathbf{Q}^T\mathbf{Q}\hat{\mathbf{x}} = \hat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$ leading to a very important result (think about how to interpret the coefficients when for a given coding scheme represented by an orthogonal matrix).

We built orthonormal vectors using **Gram-Schmidt**. Assume $\mathbf{a}$ and $\mathbf{b}$ are independent. We want to get orthogonal vectors $\mathbf{v}$ and $\mathbf{w}$ from them, and make them orthonormal by dividing by their respective lengths. Just let $\mathbf{a} = \mathbf{v}$ and $\mathbf{w} = \mathbf{b} - \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$, so we should have $\mathbf{v}^T\mathbf{w} = \mathbf{a}^T(\mathbf{b} - \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a})$ which is obviously zero. So now if you threw in a third vector $\mathbf{c}$, then we let $\mathbf{y} = \mathbf{c} - \frac{\mathbf{a}^T\mathbf{c}}{\mathbf{a}^T\mathbf{a}}\mathbf{a} - \frac{\mathbf{b}^T\mathbf{c}}{\mathbf{b}^T\mathbf{b}}\mathbf{b}$ and we now have three orthogonal vectors. So now we can write the matrix $\mathbf{Q}$ whose columns are $\mathbf{v}/||\mathbf{v}||$, $\mathbf{w}/||\mathbf{w}||$, and $\mathbf{y}/||\mathbf{y}||$. We put the three orthonormal vectors together in the matrix $\mathbf{Q}$ to emphasize the fact that $\mathbf{Q}$ and $\mathbf{A} = (\mathbf{a}\ \mathbf{b}\ \mathbf{c})$ have the same column space. So we can summarize Gram-Schmidt into $\mathbf{A} = \mathbf{Q}\mathbf{R}$. Because of the way Gram-Schmidt works, $\mathbf{R}$ turns out to be an upper-triangular matrix. So we have now learned about two famous factorizations, $\mathbf{A} = \mathbf{L}\mathbf{U}$ and now $\mathbf{A} = \mathbf{Q}\mathbf{R}$.

1.18   Properties of Determinants

Three properties define the determinant: (1) $\det(\mathbf{I}) = 1$, (2) exchanging rows reverses the sign of the determinant, so the determinant of permutation matrices is either 1 (even number of row exchanges) or -1 (odd number of row exchanges), (3) the determinant is a linear function **for each row**, e.g.

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ and } \begin{vmatrix} a+s & b+r \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} s & r \\ c & d \end{vmatrix}.$$

Note that we are not saying $\det(\mathbf{A} + \mathbf{B}) = \det(\mathbf{A}) + \det(\mathbf{B})$. From these three properties, we want to learn all there is about determinants. For example, if $\mathbf{A}$ has two equal rows then $\det(\mathbf{A}) = 0$ derives from property (2). Subtract a multiple of some row from another row and the determinant stays the same, which derives from property (2) and (3) combined. A row of zeros leads to a determinant of zero. The determinant of a triangular matrix is just the product of the pivots. This tells me that the quickest way to find a determinant is by doing row reduction first (where we have to factor out a -1 every time we do a row exchange). The reason for this last property is because a triangular matrix can be turned into a diagonal matrix $\mathbf{D}$ using further row reduction and $\det(\mathbf{D}) = d_1 d_2 \cdots d_n \det(\mathbf{I})$. We can use this last property to conclude that when $\mathbf{A}$ is invertible it has all its pivots (non-zero) and hence its determinant is non-zero. We can also use this property and row reduction to show that $\det(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) = ad - bc$. We also have $\det(\mathbf{A}\mathbf{B}) = \det(\mathbf{A})\det(\mathbf{B})$ which tells me that $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$. (Incidentally you can use this to show that the inverse of a diagonal matrix is obtained by taking the reciprocals of its diagonal elements.) We also have $\det(\mathbf{A}^2) = (\det(\mathbf{A}))^2$, $\det(2\mathbf{A}) = 2^n \det(\mathbf{A})$, and $\det(\mathbf{A}^T) = \det(\mathbf{A})$, which leads to saying that everything we said about rows can now also apply to columns of the matrix whose determinant we're taking. One key fact that determinants bring with them is the concept of **odd and even permutations**, e.g. if you can obtain some matrix after 7 row exchanges then you can only obtain that matrix after an odd number of exchanges.

1.19   Determinant Formulas and Cofactors

An alternative way of finding $\det(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$ is by writing it in the following form, noting that matrices with a row (or column) of zero have zero determinant. Using properties (2) and (3) from the previous section, I can then find the determinant of the remaining matrices.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = ad - bc$$

Imagine doing the same to a 3×3 matrix, i.e. splitting it into 27 summands. But many of those pieces will have zero determinant, so I want to know which won't, and call them survivors. You'll notice that the survivors of a 3×3 matrix are the matrix containing the diagonal elements and the following permutations of it:

$$\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix}, \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix}, \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix}, \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix}, \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix}, \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix}$$

This is a more accurate image than the one in which we go down the diagonals and switch signs (that only works for 3×3 matrices). So for an $n \times n$ matrix, we can write its determinant as the sum of determinants of $n!$ matrices. But now we need a big formula, and that formula is

$$\det(\mathbf{A}) = \sum_{n! \text{ terms}} \pm a_{1\alpha} a_{2\beta} \cdot a_{n\omega}$$

where $(\alpha, \beta, \ldots, \omega)$ is some permutation of $(1, 2, \ldots, n)$. The $\pm$ depends on whether the permutation is odd or even. We can see from the above formula for example why the determinant of $\mathbf{I}$ is just 1 or why the determinant of $\mathbf{A}$ and $\mathbf{A}^T$ are the same. We can use this big formula to find the determinant of the matrix given below

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 1 & \boxed{1} \\ 0 & 1 & \boxed{1} & 0 \\ 1 & \boxed{1} & 0 & 0 \\ \boxed{1} & 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 0 & \boxed{1} & 1 \\ 0 & \boxed{1} & 1 & 0 \\ \boxed{1} & 1 & 0 & 0 \\ 1 & 0 & 0 & \boxed{1} \end{pmatrix}$$

The first choice represents the permutation $(4, 3, 2, 1)$, which is even and has a determinant of 1, and the second choice represents the permutation $(3, 2, 1, 4)$ which is odd and has a determinant of -1. All other terms have determinant zero. So the determinant of the matrix is just zero. The least inventive way of thinking about the determinant is in terms of **cofactors** of a_{ij} which is the determinant of the matrix whose ith row and jth column have been deleted, call it $\bar{\mathbf{A}}_{ij}$, so $C_{ij} = \det(\bar{\mathbf{A}}_{ij})$. So the cofactor formula is

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\bar{\mathbf{A}}_{ij})$$

for all $i = 1, 2, \ldots, n$ (we could have expanded along a column instead). The cofactor method is a quick way to find determinants if you combine it with row reduction. Try the cofactor method with a 2×2 matrix to ensure it works.

1.20   Cramer's Rule, Inverse Matrix, and Volume

Consider the beautiful identity $\mathbf{A}\mathbf{C}^T = \det(\mathbf{A})\mathbf{I}$ where $\mathbf{C}$ is the **cofactor matrix**, which is just the cofactor formula for computing the determinant, because the ith row of $\mathbf{A}$ times the ith column of $\mathbf{C}^T$ gives me $\det(\mathbf{A})$ and the ith row of $\mathbf{A}$ times any other column of $\mathbf{C}^T$ gives me zero. The latter is not obvious at first. The reason that second case is always zero is because for example $\mathbf{A}_i$ times $\mathbf{C}_j$ for $i \neq j$ is like taking the determinant of a matrix whose ith and jth rows are identical. The identity gives us the formula $\mathbf{A}^{-1} = (\det(\mathbf{A}))^{-1} \mathbf{C}^T$ which you can check using the 2×2 matrix and sure enough it works.

Going back to $\mathbf{A}\mathbf{x} = \mathbf{b}$ gives me $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = (\det(\mathbf{A}))^{-1}\mathbf{C}^T\mathbf{b}$. This leads me to **Cramer's rule**: We can write $x_i = \det(\mathbf{B}_i)/\det(\mathbf{A})$, where $\mathbf{B}_i$ is the matrix $\mathbf{A}$ whose ith column was replaced with $\mathbf{b}$. This should be obvious from the multiplication $\mathbf{C}^T\mathbf{b}$. Note how Cramer's rule is disastrous in terms of time-consumption. Packages use the reduced echelon method to find inverses and row reduction with back substitution to find $\mathbf{x}$. It turns out that $|\det(\mathbf{A})|$ is the volume of the parallelepiped created by the columns of $\mathbf{A}$. This is obvious when $\mathbf{A} = \mathbf{I}$. The orthogonal matrix $\mathbf{Q}$ can thus be thought of the cube corresponding to $\mathbf{I}$ just rotated in some direction. The way to show that $\det(\mathbf{Q}) = 1$ is by using the fact that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. You can use the geometrical representation of the determinant to gain another insight into the three properties of the determinant: Properties (1) and (2) are follow easily; it's obvious that multiplying one row (or column) by a constant does the same to the volume. The last property to check is that of linearity in one row. We'll leave that one to you. Here's a little side note: If you have a triangle with the coordinates given by (x_i, y_i) $i = 1, 2, 3$, you can find the area of the triangle by solving $A = \frac{1}{2}\det(\mathbf{T})$ where $\mathbf{T}$ is the matrix with the rows $(x_i, y_i, 1)$.

1.21   Eigenvalues and Eigenvectors

We're looking for the particular vectors $\mathbf{x}$ now (called **eigenvectors**) for which $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some real or complex λ, called **eigenvalue**. Obviously if $\mathbf{A}$ is singular $\lambda = 0$ is overdue. Let's first ask ourselves what the eigenvectors and eigenvalues of the projection matrix would be. Since $\mathbf{b}$ and $\mathbf{P}\mathbf{b}$ are usually in different directions, $\mathbf{b}$ is usually not an eigenvector, unless of course $\mathbf{b}$ was in the plane spanned by the projection matrix in the first place, in which case we have $\lambda = 1$. So imagine working in 3D, we take two linearly independent vectors from the projection plane as eigenvectors. We now ask

if there is a third one that is not on the plane, because then we would have a basis for 3D. The answer is any $\mathbf{x}$ that is perpendicular to the plane, and we have $\lambda = 0$. So for a projection matrix the eigenvalues are 1 and 0. Consider now the permutation matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Two of its eigenvalues are $(1,1)^T$ and $(-1,1)^T$ (they are not perpendicular by accident) with respective eigenvalues 1 and -1. It turns out $n \times n$ matrices will have n eigenvalues, and sometimes for large matrices are not easy to find. But one neat property is that the trace of a matrix $\mathbf{A}$ is equal to the sum of its eigenvalues. Go back to the last example to see how that's the case.

It's easy to see that the solution $\mathbf{x}$ can be derived from the fact that $(\mathbf{A} - \lambda \mathbf{I})x = 0$, i.e. $\mathbf{A} - \lambda \mathbf{I}$ is singular or $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$. We call this last equation the **characteristic equation**. So we begin by finding λ and then use elimination to find the eigenvectors $\mathbf{x}$. Let $\mathbf{A} = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right)$ and you get eigenvalues 2 and 4 which give eigenvectors $(1,1)^T$ and $(-1,1)^T$ (or of course any multiple of them). Note how $\mathbf{P} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\mathbf{A} = \left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right)$ have the same eigenvectors and their eigenvalues for $\mathbf{A}$ are just the eigenvalues of $\mathbf{P}$ with 3 added to each. Furthermore, notice how we have $\mathbf{A} = 3\mathbf{I} + \mathbf{P}$. This makes sense given the new equation $(\mathbf{P} + 3\mathbf{I})\mathbf{x} = \lambda \mathbf{x}$. Things don't work out so great when you add two matrices $\mathbf{A}$ and $\mathbf{B}$. It does not follow that you can just add up their respective eigenvalues to get the eigenvalues of $\mathbf{A} + \mathbf{B}$. The reason is because their eigenvectors might differ, otherwise we could write $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$ and $\mathbf{B}\mathbf{x} = \alpha \mathbf{x}$ then $(\mathbf{A} + \mathbf{B})\mathbf{x} = (\lambda + \alpha)\mathbf{x}$. Now consider the (orthogonal) 90-deg. rotation matrix $\mathbf{Q} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ (just the sines and cosines of $\pi/2$). We already know $\lambda_1 = -\lambda_2$ and $\det(\mathbf{Q}) = 1 = \lambda_1 \lambda_2$. We expect something to go wrong, because no vector can come out parallel to itself after rotating it by 90 degrees. Alternatively, we can look at $\det(\mathbf{Q} - \lambda \mathbf{I}) = \lambda^2 + 1$, whose roots are complex, i.e. $\lambda = \pm i$. Note how the two eigenvalues are complex conjugates of each other. **The more we move away from symmetric matrices, the likelier we are to get imaginary eigenvalues.** In this specific case $\mathbf{Q}$ is **antisymmetric**, i.e. $\mathbf{Q}^T = -\mathbf{Q}$, so we have an extreme case. Here's another scenario: suppose we have the matrix $\mathbf{A} = \left(\begin{smallmatrix} 3 & 1 \\ 0 & 3 \end{smallmatrix}\right)$. Because $\mathbf{A}$ is triangular, then its eigenvalues are just the diagonal values (you can check it using the characteristic equation). The problem with this matrix is its eigenvectors. From $\mathbf{A} - \lambda \mathbf{I} = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ we have $(1,0)^T$ as one eigenvector. The problem is that I can't find another eigenvector, i.e. one that independent from the first one. So repeated eigenvalues open the possibility of a shortage of eigenvectors.

1.22 Diagonalization and Powers of A

Once we found the eigenvalues and eigenvectors of a matrix $\mathbf{A}$ we can proceed to diagonalize $\mathbf{A}$ by writing it in the form $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{D}$ where $\mathbf{S}$ is the matrix of eigenvectors (in the columns). Suppose I have n independent eigenvectors. Then the product $\mathbf{A}\mathbf{S}$ is just equal to the matrix whose columns are $\lambda_i \mathbf{x}_i$, $i = 1, 2, \ldots, n$. So I can represent the multiplication with $\mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{D}$ where $\mathbf{D}$ is just a diagonal matrix containing the eigenvalues. The fact that the eigenvectors are independent (which is true for most matrices) allows us now to invert $\mathbf{S}$ to get $\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ or $\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$. Let's say we wanted to know the eigenvalues and eigenvectors of $\mathbf{A}^2$ when $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$. A simple calculation shows that $\mathbf{A}^2 \mathbf{x} = \lambda \mathbf{A}\mathbf{x} = \lambda^2 \mathbf{x}$. But I can also use the diagonalization formula to show this: $\mathbf{A}^2 = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}\mathbf{S}\mathbf{D}\mathbf{S}^{-1} = \mathbf{S}\mathbf{D}^2\mathbf{S}^{-1}$. So the eigenvalues of $\mathbf{A}^k$ are just the eigenvalues of $\mathbf{A}$ raised to the k. So eigenvalues and eigenvectors are a great way to understand and calculate the powers of a matrix. An interesting byproduct of the above result is that $\mathbf{A}^k \to 0$ as $k \to \infty$ if and only if $|\lambda_i| < 1$ for all $i = 1, 2, \ldots, n$ (once again assuming independent eigenvectors).

A matrix $\mathbf{A}$ is sure to have n independent eigenvectors (and be **diagonalizable**) if all its eigenvalues are distinct. When the eigenvalues are not distinct, we may or may not have independent eigenvectors. For example, the eigenvalues of $\mathbf{I}$ are all 1, but there's no shortage of eigenvectors. So a diagonal matrix already has its eigenvalues sitting there. Triangular matrices are more problematic. Consider now the system of **first order difference equations** given by $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$, starting with $\mathbf{u}_0$. It follows that $\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0$. To really solve this we start by writing $\mathbf{u}_0$ as a combination of the eigenvectors $c_1 \mathbf{x}_1 + \ldots + c_n \mathbf{x}_n$. Notice how $\mathbf{A}\mathbf{u}_0 = c_1 \lambda_1 \mathbf{x}_1 + \ldots + c_n \lambda_n \mathbf{x}_n$ and $\mathbf{A}^k \mathbf{u}_0$ just has the eigenvalues on the right side raised to the power of k. In short, $\mathbf{A}^k \mathbf{u}_0 = \mathbf{D}^k \mathbf{S}\mathbf{c}$. We'll use the **Fibonacci sequence** as an example, given by $0, 1, 1, 2, 3, 5, \ldots$. So $F_{n+2} = F_{n+1} + F_n$. We are also interested in knowing how fast this sequence increases, and the answer lies in eigenvalues. Let's first artificially make this into a system by adding the equation $F_{k+1} = F_{k+1}$. Now we have $\mathbf{u}_k = (F_{k+1}, F_k)^T$ and $\mathbf{u}_{k+1} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)\mathbf{u}_k$. The eigenvalues are given by $\frac{1}{2}(1 \pm \sqrt{5})$. One eigenvalue is greater than 1 and one is less than 1. So we have for example $F_100 \approx c_1(\frac{1+\sqrt{5}}{2})^100$ for some $c_1 \in \mathbb{R}$, which gives us an idea about the growth rate of the sequence. Furthermore, to complete the problem we now find the eigenvectors which turn out to be $\mathbf{x}_1 = (\lambda_1, 1)^T$ and $\mathbf{x}_2 = (\lambda_2, 1)^T$. Given $\mathbf{u}_0 = (1, 0)^T$ I can now find c_1 and c_2 such that $c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 = \mathbf{u}_0$.

1.23 Differential Equations and exp(At)

Consider the differential equation given by $\frac{du_1}{dt} = -u_1 + 2u_2$ and $\frac{du_2}{dt} = u_1 - 2u_2$ with the initial condition $u(0) = (1,0)^T$. Since $\mathbf{A} = \left(\begin{smallmatrix} -1 & 2 \\ 1 & -2 \end{smallmatrix}\right)$ is singular I expect $\lambda_1 = 0$ which makes $\lambda_2 = -3$. So $\mathbf{x}_1$ is in the nullspace of $\mathbf{A}$ and we have $\mathbf{x}_1 = (2,1)^T$ and $\mathbf{x}_2$ is in the nullspace of $\mathbf{A} + 3\mathbf{I} = \left(\begin{smallmatrix} 2 & 2 \\ 1 & 1 \end{smallmatrix}\right)$ so we have $\mathbf{x}_2 = (1,-1)^T$. The solution will be $\mathbf{u}(t) = c_1 e^{\lambda_1 t}\mathbf{x}_1 + c_2 e^{\lambda_2 t}\mathbf{x}_2$ (contrast that to the solution to the difference equation $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$, which was $c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2$) and I can see that I already know a lot about the kind of solution I have just from looking at my eigenvalues; the fact that one eigenvalue is zero tells me I have some sort of steady state and the fact that the other one is negative says that as t increases it becomes less and less relevant. Solving the above equation for $t = 0$ gives me $c_1 = 1/3$ and $c_2 = 1/3$, which derive from the equation $\mathbf{S}\mathbf{c} = \mathbf{u}(0)$ where $\mathbf{S}$ is the matrix whose columns are the eigenvectors. So the steady state $\mathbf{u}(\infty) = \frac{1}{3}(2,1)^T$. In general we have **stability**, i.e. $\mathbf{u}(t) \to 0$ which requires negative eigenvalues (or negative real parts for when λ is complex). We have a **steady state** when $\lambda_1 = 0$ and λ_2 has negative real part. We blow up when either eigenvalues have positive real parts. Both eigenvalues of a 2×2 matrix $\mathbf{A}$ are negative when trace$(\mathbf{A}) < 0$ and det$(\mathbf{A}) > 0$.

Let's now write the solution to $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ in terms of $\mathbf{S}$ and $\mathbf{D}$ where $\mathbf{u} = \mathbf{S}\mathbf{v}$. The equation becomes $\mathbf{S}\frac{d\mathbf{v}}{dt} = \mathbf{A}\mathbf{S}\mathbf{v}$ or $\frac{d\mathbf{v}}{dt} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{v} = \mathbf{D}\mathbf{v}$. So we get $\mathbf{v}(t) = e^{\mathbf{D}t}\mathbf{v}(0)$ and hence $\mathbf{u}(t) = \mathbf{S}e^{\mathbf{D}t}\mathbf{S}^{-1}\mathbf{u}(0)$. So now we have to define what we mean by e raised to some matrix. We have $e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{1}{2}(\mathbf{A}t)^2 + \ldots + \frac{1}{n!}(\mathbf{A}t)^n$. (Incidentally we also have $(\mathbf{I} - \mathbf{A}t)^{-1} = \mathbf{I} + \mathbf{A}t + (\mathbf{A}t)^2 + \ldots$, which will blow up unless the eigenvalues of $\mathbf{A}t$ are below 1.) We can use this definition to show that we get $\mathbf{u}(t) = \mathbf{S}e^{\mathbf{D}t}\mathbf{S}^{-1}\mathbf{u}(0)$ is equal to $e^{\mathbf{A}t}\mathbf{u}(0)$, by writing

$$e^{\mathbf{A}t} = \mathbf{S}\mathbf{S}^{-1} + \mathbf{S}\mathbf{D}\mathbf{S}^{-1}t + \frac{1}{2}\mathbf{S}\mathbf{D}^2\mathbf{S}^{-1}t^2 + \ldots = \mathbf{S}e^{\mathbf{D}t}\mathbf{S}^{-1}.$$

But unlike the original formula which always works, note that the above formula works only as long as $\mathbf{A}$ is diagonalizable. Finally, we're left to defining the exponential of a diagonal matrix, $e^{\mathbf{D}t}$, and we define that as the diagonal matrix whose diagonal elements are $e^{\lambda_i t}$.

For a second order differential equation $y'' + by' + Ky = 0$ can be solved using a first order equation by letting $\mathbf{u} = (y', y)^T$ and hence $\mathbf{u}' = \left(\begin{smallmatrix} y'' \\ y' \end{smallmatrix}\right) = \left(\begin{smallmatrix} -b & -K \\ 1 & 0 \end{smallmatrix}\right)\mathbf{u}$. We can extend this to any nth order differential equation. This is amazing.

1.24 Markov Matrices; Fourier Series

We're back to the equation $\mathbf{u}_{k+1} = \mathbf{A}\mathbf{u}_k$ except now $\mathbf{A}$ is a Markov matrix. **Markov matrices** are matrices whose entries are all positive or zero and whose columns sum to 1. The second property ensures that a Markov matrix will have an eigenvalue of $\lambda_1 = 1$ (signaling a steady state). All other eigenvalues are less than 1 in absolute value. Furthermore, the eigenvector $\mathbf{x}_1$ has only positive elements. The matrix $\mathbf{A} - \mathbf{I}$ has columns that add up to zero and is singular because the rows are dependent, i.e. the vector $(1, 1, \ldots, 1)^T$ is in the left nullspace of $\mathbf{A}$. It follows that $\mathbf{x}_1$ is in the nullspace of $\mathbf{A}$. **The eigenvalues of $\mathbf{A}$ and $\mathbf{A}^T$ are the same.** This last piece follows because $\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{A} - \lambda\mathbf{I})^T = \det(\mathbf{A}^T - \lambda\mathbf{I}) = 0$. Here's an example: Let $\mathbf{A} = \left(\begin{smallmatrix} .9 & .2 \\ .1 & .8 \end{smallmatrix}\right)$. We have $\lambda_1 = 1$ and $\lambda_2 = .7$ which give $\mathbf{x}_1 = (2, 1)^T$ (so in the steady state we have $2/3$ and $1/3$) and $\mathbf{x}_2 = (-1, 1)^T$. So we have $\mathbf{u}_k = c_1(1^k)\left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right) + c_2(.7^k)\left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$. We can find c_1 and c_2 given some initial condition.

Let's now talk about projections with orthonormal bases $\mathbf{q}_1, \ldots, \mathbf{q}_n$. For some vector $\mathbf{v} = x_1\mathbf{q}_1 + \ldots + x_n\mathbf{q}_n$ we want to find a way to isolate x_1: Take the dot product of everything with $\mathbf{q}_1$ and get $\mathbf{q}_1^T\mathbf{v} = x_1$. In matrix notation we have $\mathbf{v} = \mathbf{Q}\mathbf{x}$ and hence $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{v} = \mathbf{Q}^T\mathbf{v}$. This is what **Fourier series** are built on. Fourier series look like $f(x) = a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + b_2\cos 3x + \ldots$. Note that we are now working with an **infinite dimension vector space**. We still need to define what the dot product of functions are. For functions f and g we have $<f, g> = \int f(x)g(x)dx$ with the integration occurring on the interval $[0, 2\pi]$. It's easy to check that for example $\int_0^{2\pi} \sin x \cos x \, dx = 0$ and the same is true for all the other trig functions. Now we need to find the constants. We begin by setting a_0 equal to the average of f. To find a_1 we multiply the whole thing by $\cos x$ and integrate:

$$\int_0^{2\pi} f(x)\cos x \, dx = a_1 \int_0^{2\pi} (\cos x)^2 \, dx = a_1\pi$$

I can find all the other a_i and b_i in a similar fashion.

1.25 Symmetric Matrices and Positive Definiteness

Symmetric matrices are the most important class of matrices. It turns out that real symmetric matrices have real eigenvalues and their eigenvectors are (can be chosen to be) orthogonal and can hence be made orthonormal (proofs are omitted). The equation $\mathbf{A} = \mathbf{SDS}^{-1}$ ($\mathbf{D}$ is the diagonal matrix containing the eigenvalues) in the case of a symmetric $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{QDQ}^T$ where $\mathbf{Q}$ is an orthogonal matrix. This amazing factorization is one of the famous theorems of linear algebra. We can expand the factorization as follows

$$\mathbf{A} = \mathbf{QDQ}^T = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{pmatrix} \mathbf{D} \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{pmatrix} = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \ldots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$$

where $\mathbf{q}_i \mathbf{q}_i^T$ is a projection matrix. So **every symmetric matrix is a combination of mutually perpendicular projection matrices.** Now let's find a way to determine if the real eigenvalues of a symmetric matrix are positive or negative. It turns out the signs of the pivots are the same as the sign of the eigenvalues. So positive pivots have positive eigenvalues and negative pivots negative ones. Remember that the product of the pivots give us the determinant, which in turns equals the product of the eigenvalues.

If symmetric matrices are good, wait till you hear about symmetric **positive definite** matrices, i.e. matrices with real and positive eigenvalues (and hence real and positive pivots). For a symmetric matrix to be positive definite, all its sub-determinants must be positive.

1.26 Complex Matrices; Fast Fourier Transform

Let $\mathbf{z} \in \mathbb{C}^n$, then the length of $\mathbf{z}$ is not given by $\mathbf{z}^T\mathbf{z}$ anymore; instead we use $\bar{\mathbf{z}}^T\mathbf{z} = |z_1|^2 + \ldots + |z_n|^2$, where $\bar{\mathbf{z}}$ is the complex conjugate of $\mathbf{z}$. This is just an extension of $(a+bi)(a-bi) = a^2 + b^2$. We use the notation $\bar{\mathbf{z}}^T\mathbf{z} = \mathbf{z}^H\mathbf{z}$ to combine both actions where the H stands for **Hermitian**. Similarly, if $\mathbf{A}$ is a **complex matrix**, then it is a symmetric matrix when $\bar{\mathbf{A}}^T = \mathbf{A}$, so for example the matrix $\begin{pmatrix} 2 & 3+i \\ 3-i & 7 \end{pmatrix}$ is a symmetric complex matrix, also known as a **Hermitian matrix**. To check that the complex vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ are orthonormal, then $\bar{\mathbf{q}}_i^T\mathbf{q}_j$ must be 1 when $i = j$ and 0 otherwise. In other words I must have $\bar{\mathbf{Q}}^T\mathbf{Q} = \mathbf{Q}^H\mathbf{Q} = \mathbf{I}$. When dealing with complex matrices, we usually change the word symmetric with Hermitian, and the word orthogonal matrix (orthonormal columns) with the word **unitary**. The most famous complex matrix (that's also unitary) is the Fourier matrix, given by

$$\mathbf{F}_n = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W & W^2 & \cdots & W^{n-1} \\ 1 & W^2 & W^4 & \cdots & W^{2(n-1)} \\ \vdots & & & & \vdots \\ 1 & W^{(n-1)} & W^{2(n-1)} & \cdots & W^{(n-1)^2} \end{pmatrix}$$

such that the ith row and jth column of $\mathbf{F}_n$ is given by W^{ij} with $i, j = 0, 1, \ldots, n-1$. W is given by $e^{i2\pi/n} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$ and note that its nth power is 1, so for a given n, W is the complex number that lies on the unit circle when starting at zero I go $2\pi/n$ counterclockwise, W^2 is when I go another $2\pi/n$ counterclockwise, and by the time I'm at W^n I'm back to where I started. For example, when $n = 4$ we have $W^4 = 1$ and $W = e^{2\pi i/4} = i$, so the powers of W from 1 to 4 are i, -1, $-i$, and 1. The matrix $\mathbf{F}_4$ is given by

$$\mathbf{F}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

You can check that the above matrix has orthogonal columns, and by dividing the matrix by 2 we turn it into an orthogonal matrix (i.e. we make the columns orthonormal). So $\mathbf{F}_4^H\mathbf{F}_4 = \mathbf{I}$. Why is this matrix so remarkable? It has to do with how $W_{2n}^2 = W_n$, which allows me to write $\mathbf{F}_n$ for some large n in terms of smaller n's and thereby reduce the number of computations. That's why it's considered a major leap in scientific computing. We'll leave that for you to discover.

1.27 Positive Definite Matrices and Minima

We're back to real matrices now. A 2×2 matrix $\mathbf{A} = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ is positive definite (positive definite matrices are always symmetric) if (1) it has positive eigenvalues, or (2) it has positive sub-determinants, i.e. $a > 0$ and $ac - b^2 > 0$, or (3) it has positive pivots, i.e. $a > 0$ and $(ac - b^2)/a > 0$, or (4) if $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ with equality only for $\mathbf{x} = 0$. The last definition is the more conventional one; all the other definitions are more like tests. On the other hand, a matrix such as $\left(\begin{smallmatrix} 2 & 6 \\ 6 & 20 \end{smallmatrix}\right)$ is called a **positive semi-definite** matrix; its determinant is zero and one of its eigenvalues is zero (while the other one is positive), and it only has one pivot. The multiplication $\mathbf{x}^T \mathbf{A} \mathbf{x}$ gives $2x_1^2 + 12x_1 x_2 + 20x_2^2$, called a **quadratic form**. I want to know if for all $x_1, x_2 \in \mathbb{R}$ that quadratic form is positive or not. The matrix $\left(\begin{smallmatrix} 2 & 6 \\ 6 & 7 \end{smallmatrix}\right)$ is not positive definite and it corresponding graph $(x, y, f(x, y))$ where $f(x, y) = \mathbf{x}^T \mathbf{A} \mathbf{x} = 2x^2 + 12xy + 7y^2$ goes up is some directions but down in others. In other words, the graph has a saddle point (second derivative test fails). We will see later that the perfect directions to look are the eigenvectors' directions. The graph of a positive definite matrix on the other hand can only go up as we move away from the origin. To generalize, for the function $f(x_1, \ldots, x_n)$ to have a minimum at some point, the matrix of its second derivatives must be positive definite. In other words, the diagonal elements in the matrix below must overwhelm the nondiagonal elements.

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

If $2x_1^2 + 12x_1 x_2 + 20x_2^2$ is positive definite, then I should be able to complete the squares which gives $2(x + 3y)^2 + 2y^2$. So for any fixed $z = f(x, y)$ I get a cross section that is an ellipse. Notice how the multipliers in the equation we got from completing the square, 2, 3, and 2, correspond to the first pivot, the multiplier that gives me the second pivot when I do row reduction, and the second pivot respectively. In other words, $\left(\begin{smallmatrix} 2 & 6 \\ 6 & 20 \end{smallmatrix}\right) \rightarrow \left(\begin{smallmatrix} 2 & 6 \\ 0 & 2 \end{smallmatrix}\right)$ is what I get when I subtract three times row 1 from row 2. I can now see why I needed positive pivots. This fantastic result allows me to extend my results to $n \times n$ matrices, since I know how to perform row reduction on them. Consider the 3×3 matrix given by

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

which has sub-determinants 2, 3, and 4. Since the product of pivots must give me the determinants, then the pivots are 2, $\frac{3}{2}$ and $\frac{4}{3}$ (because $2(\frac{3}{2}) = 3$ which is the second sub-determinant and $3(\frac{4}{3}) = 4$, the determinant). The equation $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is given by $f(x_1, x_2, x_3) = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1 x_2 - 2x_2 x_3$ which we know is always positive and its cross section for a fixed $z = f(x_1, x_2, x_3)$ is an ellipsoid. The ellipsoid in this case will have 3 axes, (the same way an ellipse has 2 axes), which are given by the eigenvalues of the matrix $\mathbf{A}$. The direction of those axes are given by the eigenvectors corresponding to those eigenvalues. So I can imagine what the ellipsoid looks like by diagonalizing it into $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$ (which I can do because $\mathbf{A}$ is symmetric).

1.28 Similar Matrices and Jordan Form

We might ask where positive definite matrices come from. From the properties of positive definite matrices, I can now say that the inverse of positive definite matrices are also positive definite. The sum of positive definite matrices is also positive definite, because $\mathbf{x}^T(\mathbf{A} + \mathbf{B})\mathbf{x} > 0$ unless $\mathbf{x} = 0$. Now suppose $\mathbf{A}$ is an $m \times n$ rectangular matrix. Then we know that $\mathbf{A}^T \mathbf{A}$ is square and symmetric. So how do we know if it's positive definite (or at least positive semi-definite)? The easiest way to answer this is by studying $\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}$ which is just $(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = ||\mathbf{A}\mathbf{x}||^2$ which is zero only when the vector is zero. So **for an $m \times n$ matrix as long as** $\text{rank}(\mathbf{A}) = n$ **the matrix $\mathbf{A}^T \mathbf{A}$ is always positive definite**. In other words, the nullspace of $\mathbf{A}$ must be 0.

Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are **similar** if for some (invertible) matrix $\mathbf{M}$ we can write $\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$. For example, $\mathbf{A}$ is always similar to the matrix $\mathbf{D}$ because we have $\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$. So matrices can be partitioned into "families" with all matrices in a family being similar (and the nicest of them being the diagonal matrix $\mathbf{D}$). This means that **similar matrices have the same eigenvalues**. This is because from $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ we have $\mathbf{M}^{-1}\mathbf{A}\mathbf{M}\mathbf{M}^{-1}\mathbf{x} = \mathbf{B}\mathbf{M}^{-1} = \lambda\mathbf{M}^{-1}\mathbf{x}$. So while the eigenvalues stay the same, the eigenvectors of $\mathbf{B}$ are obtained by multiplying $\mathbf{M}^{-1}$ by the eigenvectors of $\mathbf{A}$. The results above assumed that the eigenvalues are distinct (otherwise we would not have all the eigenvectors required to diagonalize the matrix). Take now the case $\mathbf{A} = \left(\begin{smallmatrix} 4 & 0 \\ 0 & 4 \end{smallmatrix}\right)$ has eigenvalues

4 and 4. But this matrix is not similar to let's say $\mathbf{B} = \left(\begin{smallmatrix} 4 & 0 \\ 1 & 4 \end{smallmatrix}\right)$. In fact $\mathbf{A}$ is in one small family with the identity and $\mathbf{B}$ has a larger family but it is the best of its family and has a form called the **Jordan form**. Jordan basically found the best looking matrix in each family and completed diagonalization by coming as near to it as possible in cases where repeated eigenvalues exist. Finding the Jordan form of a matrix requires that you know the exact eigenvalues and the rank of the matrix. The matrix $\left(\begin{smallmatrix} 5 & 1 \\ -1 & 3 \end{smallmatrix}\right)$ is in the same family as the matrix $\mathbf{B}$ above (because they have the same trace and determinant) and hence it is not diagonalizable. You can read more on Jordan forms if you want.

1.29 Singular Value Decomposition

SVD is a central part of linear algebra and where everything in this course is generalized. It is the final and best factorization of a matrix. $\mathbf{A}$ can be any matrix, and can be written as the product of two orthogonal matrices and a diagonal one, given by $\mathbf{A} = \mathbf{U\Sigma V}^T$ where $\mathbf{\Sigma}$ is the diagonal matrix. So far we learned that when $\mathbf{A}$ is a symmetric matrix we can write $\mathbf{A} = \mathbf{SDS}^{-1}$ where $\mathbf{S}$ is the matrix of eigenvectors, and when $\mathbf{A}$ is positive definite we have $\mathbf{A} = \mathbf{QDQ}^T$. So one orthogonal matrix $\mathbf{Q}$ gets me my SVD in the (symmetric) positive definite case. **Singular value decomposition** extends this to non-symmetric matrices and tries to do is take an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in the row space of $\mathbf{A}$ (by using Gram-Schmitt for example) and have it go over to an orthogonal basis $\sigma_1 \mathbf{u}_1, \ldots, \sigma_k \mathbf{u}_k$ ($\mathbf{u}_i$'s are orthonormal) in the column space of $\mathbf{A}$ where $\sigma_i \mathbf{u}_i = \mathbf{Av}_i$ for all $i = 1, \ldots, k$. Usually, there's no reason for us to believe that just because the $\mathbf{v}$'s are orthogonal then so will the $\mathbf{u}$'s. But SVD gives us that guarantee. So we have $\mathbf{AV} = \mathbf{U\Sigma}$ where $\mathbf{V}$ is the matrix whose columns are the $\mathbf{v}_i$'s and $\mathbf{U}$ is the matrix whose columns are $\mathbf{u}_i$'s and $\mathbf{\Sigma}$ is the diagonal matrix containing σ_i as its diagonal elements. How do we deal with nullspaces? Just add the bases for the null space to $\mathbf{V}$ and add some 0's to the diagonal elements of $\mathbf{\Sigma}$. So we want obtain $\mathbf{A} = \mathbf{U\Sigma V}^{-1}$ and since $\mathbf{V}$ is orthogonal we simplify to $\mathbf{A} = \mathbf{U\Sigma V}^T$. From the last equation we get

$$\mathbf{A}^T\mathbf{A} = \mathbf{V\Sigma}^T\mathbf{U}^T\mathbf{U\Sigma V}^T = \mathbf{V\Sigma}^T\mathbf{I\Sigma V}^T = \mathbf{V}(\mathbf{\Sigma}^2)\mathbf{V}^T$$

which shows that $\mathbf{V}$ is the vector containing the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and $\sigma_1^2, \ldots, \sigma_n^2$ are its eigenvalues. So this gives me my $\mathbf{V}$ vectors. Similarly, we can show that $\mathbf{U}$ is the matrix of eigenvectors for $\mathbf{AA}^T$.

For example, let $\mathbf{A} = \left(\begin{smallmatrix} 4 & 4 \\ -3 & 3 \end{smallmatrix}\right)$ and so $\mathbf{A}^T\mathbf{A} = \left(\begin{smallmatrix} 25 & 7 \\ 7 & 25 \end{smallmatrix}\right)$. I should know by now that a matrix like this gives me two eigenvectors: $(1,1)^T$ and $(1,-1)^T$, which I normalize by dividing by $\sqrt{2}$. It quickly becomes obvious that my eigenvalues are 32 and 18. So I have

$$\mathbf{A} = \begin{pmatrix} 4 & 4 \\ -3 & 3 \end{pmatrix} = \mathbf{U} \begin{pmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Now we compute $\mathbf{AA}^T$ which by chance turns out diagonal and is given by $\left(\begin{smallmatrix} 32 & 0 \\ 0 & 18 \end{smallmatrix}\right)$ and its eigenvectors are $(1,0)^T$ and $(0,1)^T$. (Notice how the eigenvalues stayed the same when you change the order of multiplication; this is always true.) So we get $\mathbf{U} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Take an example of a singular matrix $\mathbf{A} = \left(\begin{smallmatrix} 4 & 3 \\ 8 & 6 \end{smallmatrix}\right)$. The row space of $\mathbf{A}$ is all multiples of $(4,3)^T$ and the unit vector of that row space is $\mathbf{v}_1 = (.8, .6)^T$ and $\mathbf{v}_2$ will come from the nullspace of $\mathbf{A}$. The column space of $\mathbf{A}$ is given by all multiples of $(4,8)^T$ and the unit vector of that space is $\mathbf{u}_1 = \frac{1}{\sqrt{5}}(1,2)^T$ and $\mathbf{u}_2$ will come from the left nullspace of $\mathbf{A}$. This will give us $\mathbf{U}$ and $\mathbf{V}$. We can find $\mathbf{\Sigma}$ the usual way but we already expect σ_2 to be zero. In general, let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be an orthonormal basis for the row space of $\mathbf{A}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_r$ be an orthonormal basis for its column space. Then $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ and $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ will be an orthonormal basis for $N(\mathbf{A})$ and $N(\mathbf{A}^T)$ respectively. Then we have $\mathbf{Av}_i = \sigma_i \mathbf{u}_i$.

1.30 Linear Transformations and Their Matrices

In a way we are now restarting the course and taking a coordinate-free approach by using **linear transformation**. Think for example of the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that projects any vector in the 2 dimensional plane onto a vector in 2 dimensional plane. So the vector is the input and its projection the output, and I don't need any coordinates (i.e. axes) to get the output from the input. A transformation is linear if $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and $T(c\mathbf{x}) = cT(\mathbf{x})$. These two conditions imply that $T(0) = 0$ and $T(a\mathbf{x}+b\mathbf{y}) = aT(\mathbf{x})+bT(\mathbf{y})$. For example, the transformation that shift any vector by some vector $\mathbf{v}_0$ is not linear, neither is the transformation that takes any vector and produces its length. Projection and rotation on the other hand are both linear transformations, so is $T(\mathbf{v}) = \mathbf{Av}$ for some matrix $\mathbf{A}$; if $\mathbf{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ then my transformation consists of flipping everything over the x-axis. Our ultimate goal is to find the matrix that corresponds to

some linear transformation. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an input basis, then $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ is a basis for the output space, because for every $\mathbf{v} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n$ I have $T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + \ldots + c_n T(\mathbf{v}_n)$. Coordinates come from a basis, because every vector can be written as the sum of multiples of vectors in the basis and those multiples are the coordinates. We now want to construct the matrix $\mathbf{A}$ that represents the linear transformation T. We first have to choose a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ for the input and a basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ for the output. Now, I can express an input vector in terms of its coordinates and find a matrix $\mathbf{A}$ that when multiplied by gives me the coordinates of the output vector. The correct choice of basis can make the job of finding $\mathbf{A}$ significantly easier. For example, for the matrix corresponding to projection on a line in $\mathbb{R}^2$ choosing the eigenvector basis gives me a unit vector on the projection line $\mathbf{v}_1$ and a unit vector perpendicular to it $\mathbf{v}_2$. Let $\mathbf{w}_1 = \mathbf{v}_1$ and $\mathbf{w}_2 = \mathbf{v}_2$ then projection will kill $\mathbf{w}_2$ and leave $\mathbf{w}_1$, which corresponds to the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ because $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} c_1 \\ 0 \end{smallmatrix}\right)$. So the eigenvector basis leads to the diagonal matrix of eigenvalues. If we use the standard basis in $\mathbb{R}^2$, $\mathbf{e}_1$ and $\mathbf{e}_2$, then the projection matrix that projects on the 45-deg. line (as we showed earlier) is $\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}} = \left(\begin{smallmatrix} .5 & .5 \\ .5 & .5 \end{smallmatrix}\right)$ but this matrix is not as nice as the last matrix. In general, to find the matrix $\mathbf{A}$ given the bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \ldots, \mathbf{w}_m$ and starting with $\mathbf{v}_1$ write $T(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + \ldots + a_{m1}\mathbf{w}_m$ and the vector $(a_{11}, \ldots, a_{m1})^T$ becomes the first column of $\mathbf{A}$, and we repeat this for $\mathbf{v}_2$ through $\mathbf{v}_n$. Consider now the linear transformation $T = \frac{d}{dx}$ with the input space being $c_0 + c_1 x + c_2 x^2$ (thus the basis is 1,x, and x^2) and the output space being $c_1 + 2c_2 x$ (with basis 1 and x), and we have $\mathbf{A} = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right)$. Therefore the derivative is a linear transformation.

1.31   Change of Basis; Image Compression

A change of basis is something that is commonly done in applications of linear algebra. It is for example widely used in image compression. Imagine a grayscale picture that is 512×512 pixels where each pixel x_i can take a value between 0 and 255 (in 8 bits). The information for the whole image can thus be stored in and represented by the vector $\mathbf{x} \in \mathbb{R}^n$ where in this case $n = (512)^2$. A common compression scheme is called JPEG, and it is essentially a change of basis. Compression takes advantage of the fact that many of the values neighboring in the vector $\mathbf{x}$ are very close to each other. A standard basis for the space of $\mathbf{x}$ would consist of the standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. But we need to find a better basis and we can take the following orthogonal basis (which is known as the **Fourier basis** and it should be obvious why):

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ -1 \\ 1 \\ \vdots \\ 1 \\ -1 \end{pmatrix}$$

Note how the first vector corresponds to a black (or white, you pick) picture. The next one is half black half white and by the time you reach the last vector you have a checkers board. Think of it as a basis that starts with low-frequency vectors and moves to higher and higher frequency ones. JPEG usually breaks the image into blocks of 8 by 8 pixels and changes the basis on each of those blocks separately. So in comes a signal in the form of a vector $\mathbf{x} \in \mathbb{R}^{64 \times 64}$, and we change basis and obtain the coefficients $\mathbf{c}_i$ for $\mathbf{x}$. Now comes the actual compression step, which produces $\hat{c}_i$ and we can write $\hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$ but now most of the last $\mathbf{c}_i$'s are just zero.

The competition to a Fourier basis is called **wavelets**. It is expected that in the future people will come up with even better bases. A good basis should be fast (such as FFT, or the fast wavelet transform FWT), orthogonal, and only a few basis vectors are enough to reproduce the image. Let's say a linear transformation is represented by matrix $\mathbf{A}$ with respect to besis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and matrix $\mathbf{B}$ with respect to basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ then the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar, i.e. $\mathbf{B} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ where $\mathbf{M}$ is the matric whose columns are the new basis vectors.