Lecture 4-14

Today we present a couple of examples of linear transformations(=linear maps) and their corresponding matrices. We begin with the very familiar setting of the plane $\mathbb{R}^2$. An easy example of a linear transformation from $\mathbb{R}^2$ to itself is rotation by θ radians counterclockwise; this sends the first unit coordinate vector $(1,0)$ to $(\cos\theta, \sin\theta)$ and $(0,1)$ to $(-\sin\theta, \cos\theta)$. Accordingly, its matrix M_θ is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Note that the matrix product $M_\theta M_{\theta'} = M_{\theta+\theta'}$; rotation by θ followed by rotation by θ' amounts to rotation by $\theta+\theta'$. A somewhat more subtle example in $\mathbb{R}^2$ is reflection by the line through the origin corresponding to the angle $\theta/2$; this sends $(1,0)$ to $(\cos\theta, \sin\theta)$ as before, but now sends $(0,1)$ to $(\sin\theta, -\cos\theta)$. Accordingly, the matrix R_θ of this reflection is $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$. It sends $(\cos\theta/2, \sin\theta/2)$ to itself but sends $(-\sin\theta/2, \cos\theta/2)$ to its negative. Note that a typical product $R_\theta R'_\theta$ is *not* equal to R_α for any α; in terms of linear transformations, the product of two reflections is not a reflection. Instead it is a rotation; as an exercise, work out the angle by which it is rotation. Rotations can also be defined in $\mathbb{R}^n$ for $n \geq 3$; for example, the rotation by θ radians counterclockwise in the xy-plane makes sense in $\mathbb{R}^3$ and has matrix $\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Here however there are many more possibilities; you can rotate in any plane containing $\vec{0}$, not just the xy-plane, and thereby fix any normal vector of this plane. A remarkable fact which we may prove later is that the composite of two rotations in $\mathbb{R}^3$, *even in different planes*, is another rotation in $\mathbb{R}^3$ in yet a third plane; this is not at all obvious.

We can also relate row operations to matrix multiplication. Specifically, call an $n \times n$ matrix *elementary* if it is obtained from the identity matrix I by applying a single row operation (either adding one multiple of a row to another or interchanging two rows). Then you can check that *if E is an elementary $m \times m$ matrix obtained from I by applying the row operation π, then the product EM is the matrix obtained from M by applying the same row operation π to M, for any $m \times n$ matrix M.* If you multiply on the right rather than the left, then row operations get replaced by column operations: if E is obtained from the $n \times n$ identity matrix I by a single *column* operation (adding one multiple of a column to another or interchanging two columns) then the product ME of an $m \times n$ matrix M and E is obtained from M by applying the same column operation to M. Note that applying a column operation to the coefficient matrix of a linear system $MX = B$, unlike applying a row operation, does change its solution set, but only by changing its variables. More precisely, if α times column j is added to column i of the matrix M in the system $MX = B$ (with $j > i$), then a solution $(x_1, \ldots, x_i, \ldots, x_j \ldots, x_n)$ (written as a column vector) to the new system gives rise to a solution $(x_1, \ldots, x_i, \ldots, x_j + \alpha x_i, x_{j+1}, \ldots, x_n)$ to the old system; similarly, if columns i and j are interchanged in M, then a solution to the new system becomes a solution to the old system $MX = B$ if its ith and jth coordinates are interchanged. In particular, the system obtained from a system $MX = B$ by applying a column operation to M is consistent if and only if $MX = B$ is consistent and it has a nonzero solution if and only if $MX = B$ does. We also see that any $m \times n$ matrix M can be brought to echelon form by multiplying it on the left by a suitable product of elementary matrices. (Note also that the two senses of "elementary matrix" are equivalent here: a square matrix is obtained from I by a single row operation if and only if it is obtained from I by a single column operation.)

A matrix M in echelon form is said to be in *reduced echelon form* if any column of it with a pivot has the pivot as its *only* nonzero entry (that is, zeroes occur above the pivot position in any column with a pivot as well as below this position) *and* only 1s occur in pivot positions. Given a matrix in echelon form we can bring it to reduced echelon form by subtracting a suitable multiples of the second row (if it has a pivot) from the first one, then a suitable multiple of the third row from the first one, and so on, until all pivots in lower rows than the first have zeroes about them in the first row, then doing the same for the second row, and so on, followed by dividing each nonzero row by the number in its pivot position. We will see an example of this and work out the consequences of being in reduced echelon form next time. For now, we mention a variation of the echelon form which applies to square *symmetric* matrices. Given an $n \times n$ symmetric matrix M, suppose that its 11th entry m_{11} is not 0. Do the usual row operations to replace all all entries in the first column of M below m_{11} by 0, but now follow these up by doing the corresponding *column* operations at the same time and in the same order. (For example, if you added 2 times row 1 to row 3, add 2 times column 1 to column 3). The effect of these additional operations is to "symmetrize" the effect of the original row operations, so that the matrix M' thus obtained is again symmetric and has zeros in its first row to the right of its first entry, just as it continues to zeros in its first column below this entry. Now do the same for the second row, assuming that current 22-entry is nonzero; having applied row operations to make all entries below the 22th entry 0, follow these up with the corresponding column operations, so that the new matrix is again symmetric and its second row and column consist entirely of zeros apart from the 22th entry.

Continue in this way with the subsequent columns, assuming that all ii entries are nonzero as required. The upshot is that you get a *diagonal* matrix D (whose only nonzero entries lie on the main diagonal, from upper left to lower right). Now we have seen that any row operationis implemented by multiplying the matrix on the left by suitable elementary matrices E_i. Every time you multiply the matrix on the left by some E_i, you multiply it on the right by the *transpose* E_i^T of E_i (obtained from the identity matrix by doing the same operation on the columns that was done on the rows from I to produce E_i), so as to do the corresponding column operation on the matrix.

Recall that when we were discussing Hessian matrices H earlier of differentiable functions at critical points, we saw that these matrices are always symmetric and the critical point is a local minimum whenever the Hessian matrix H is positive definite in the sense that whenever H is multiplied by a nonzero row vector $\vec{v}$ on the left and by that same vector on the right, written as a column vector, the resulting number is positive. It turns out that this property holds of H if and only if it holds of the diagonal matrix D produced by the above procedure; in turn this holds if and only if all diagonal entries of D are positive. We now have a practical algorithm for deciding whether a symmetric matrix is positive definite. We will return to this topic later.