Matrices and linear equations

Basic properties of matrices

In the first year you have seen how to manipulate matrices, and use them to solve systems of simultaneous equations. This handout is a brief review of some of that material.

Recall that we are taking $\mathbb{F}$ to be either $\mathbb{R}$ or $\mathbb{C}$. An $m \times n$ **matrix over** $\mathbb{F}$ is a rectangular array of elements from $\mathbb{F}$ (we call elements of $\mathbb{F}$ **scalars**) with m rows and n columns. For a matrix A, we label the element in the ith row and the jth column by a_{ij}. If $m = n$ we call A a **square matrix**; if A is an $m \times 1$ (respectively a $1 \times n$) matrix we call A a **column** (respectively **row**) **vector**. We usually denote row and column vectors by lower-case boldface letters (e.g $\mathbf{a}$) — this corresponds to underlining vectors as in the lectures. (Indeed, underlining is the traditional way in a handwritten manuscript to show a printer which parts are intended to be in boldface type.)

Given two $m \times n$ matrices A and B, we define their **sum** $A + B$ to be the $m \times n$ matrix C with $c_{ij} = a_{ij} + b_{ij}$. We can also **multiply a matrix by a scalar**: λA is the matrix C with $c_{ij} = \lambda a_{ij}$. Further, if A is an $l \times m$ matrix over $\mathbb{F}$ and B is an $m \times n$ matrix over $\mathbb{F}$, then we define their **product** AB (an $l \times n$ matrix over $\mathbb{F}$) to be the matrix C, where

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

Recall that the order of the product matters; in general $AB \neq BA$ (indeed, one of the two products might not even be defined!).

The $n \times n$ matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

is called the **identity matrix** I_n. We have for all $n \times n$ matrices A that $AI = IA = A$, and I is the only such matrix with this property.

Given A, an $n \times n$ matrix over $\mathbb{F}$, we say that an $n \times n$ matrix B is the **inverse** of A if $AB = I_n$. Such a matrix also satisfies $BA = I_n$. Given A, if such a matrix B exists then it is unique; we will denote this inverse matrix by A^{-1}. It is easy to check that $(AB)^{-1} = B^{-1}A^{-1}$. We can also take the **transpose** A^T of any $m \times n$ matrix A. This is the $n \times m$ matrix C with $c_{ij} = a_{ji}$ (i.e., the matrix obtained by interchanging the rows and columns of A). We have $(AB)^T = B^T A^T$ whenever the product AB exists. It is also possible to show that $(A^{-1})^T = (A^T)^{-1}$. The **trace** of an $n \times n$ matrix A is the sum of the elements along the diagonal; that is

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

The determinant of a (square) matrix

We can relate to each $n \times n$ matrix A over $\mathbb{F}$ an element $\det(A)$ in $\mathbb{F}$, called the **determinant** of A. There are various ways to define this, and to calculate it for any given matrix. On this sheet I will describe one possible definition; if this differs from the way that you are familiar with, check instead that you are happy with the definition that you *are* familiar with, and then concentrate on the properties of the determinant that I will later go on to describe.

A **permutation** of the integers $\{1, \ldots, n\}$ is an arrangement of these integers in some order with no repetition or eliminations. For example, the permutations of $\{1, 2, 3\}$ are

$$(1,2,3),\ (1,3,2),\ (2,1,3),\ (2,3,1),\ (3,1,2),\ (3,2,1).$$

Given such a permutation σ we set $\sigma(i)$ to be the ith element of the permutation; for example, if $\sigma = (3,1,2)$ then $\sigma(1) = 3$, $\sigma(2) = 1$, and $\sigma(3) = 2$. We denote the set of permutations of n by Σ_n.

An **inversion** occurs in a permutation σ whenever a larger integer precedes a smaller one. For example $(3,2,1)$ contains three inversions: 3 before 2, 3 before 1, and 2 before 1. We say that a permutation is **even** if it contains an even number of inversions, and **odd** otherwise; for example $(3,2,1)$ is an odd permutation. Let $l(\sigma)$ be the number of inversions in the permutation σ; in our example we have $l((3,2,1)) = 3$.

We now define the **determinant** of an $n \times n$ matrix A to be

$$\det(A) = \sum_{\sigma \in \Sigma_n} (-1)^{l(\sigma)} a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}.$$

For example, for A a 2×2 matrix we have

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

while for B a 3×3 matrix we have

$$\det(B) = b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} - b_{11}b_{23}b_{32}.$$

Calculating the determinant of a square matrix in this manner can become a lengthy procedure if the number of rows and columns is large. For this reason it is convenient to have a number of alternative ways to calculate it. Most of these rely on the fact that for any two $n \times n$ matrices A and B we have

$$\det(AB) = \det(A)\det(B).$$

For our purposes, the importance of the determinant function is

Theorem 1 *An $n \times n$ square matrix A is invertible if and only if $\det(A) \neq 0$.*

Matrices and linear equations

You have already seen how we can use matrix methods to solve systems of linear equations. Given such a system

$$\begin{array}{lllll} a_{11}x_1 & +a_{12}x_2 & +\cdots & +a_{1n}x_n & =b_1 \\ a_{21}x_1 & +a_{22}x_2 & +\cdots & +a_{2n}x_n & =b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1}x_1 & +a_{m2}x_2 & +\cdots & +a_{mn}x_n & =b_m \end{array}$$

we can rewrite it in the form

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (j = 1, 2, \ldots, m)$$

which is equivalent to the matrix equation

$$A\mathbf{x} = \mathbf{b}.$$

The matrix A is called the **coefficient matrix** of the system of equations, and the $m \times (n+1)$ matrix given by adjoining the column vector $\mathbf{b}$ to the right-hand side of A — which we will denote by $(A|\mathbf{b})$ — is called the **augmented matrix** of the system. Last year you have seen that this augmented matrix contains all the information needed to understand our system of equations.

One way to solve such a system of equations is to use elementary row operations to perform **Gaussian elimination**. Clearly the set of solutions of our equations is unchanged if we

1. Swap the positions of a pair of equations.
2. Multiply both sides of an equation by a *non-zero* scalar.
3. Add one equation to another.

These correspond to the following **elementary row operations** on the matrix $A' = (A|\mathbf{b})$:

1. Replace A' by $H^{i,j}A'$, where $H^{i,j}$ is the $n \times n$ matrix which exchanges rows i and j.
2. Replace A' by $H^i_\lambda A'$, where H^i_λ is the $n \times n$ matrix which multiplies row i by $\lambda \neq 0$.
3. Replace A' by $M^{i,j}A'$, where $M^{i,j}$ is the $n \times n$ matrix which adds rows j to row i.

[Make sure you know what the matrices $H^{i,j}$, H^i_λ and $M^{i,j}$ look like in terms of their entries; they are very simple matrices.]

Now the method of Gaussian elimination for solving systems of linear equations corresponds to *transformation of the augmented matrix by elementary row operations into Echelon form.* Recall that a matrix A is in **Echelon form** if all the entries below the first non-zero entry in each row are zero. We call the first non-zero entry of each row the **distinguished element** of that row. Then a matrix in echelon form is said to be **row reduced** if: i) all of

the distinguished elements equal 1, and ii) each is the *only* non-zero element in its respective column. For example, if A and B are given by

$$A = \begin{pmatrix} 7 & 1 & 0 & 3 & 5 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 0 & 4 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

then both A and B are in echelon form, but only B is row reduced. It is possible to show that the row reduced echelon form of a matrix is unique.

[Make sure that you are happy that Gaussian elimination corresponds to using elementary row operations to put the augmented matrix in Echelon form.]

The determinants of the matrices corresponding to elementary row operations are easy to calculate:

$$\det(H^{ij}) = -1 \qquad \det(H^i_\lambda) = \lambda \qquad \det(M^{ij}) = 1.$$

Calculating the determinant and inverse of a matrix

The determinant of a matrix can be calculated using the Laplace transformation for $\det(A)$ which, when used with elementary row operations, can greatly simplify the calculation. Given a $n \times n$ matrix A, the **minor** of a_{ij}, denoted M_{ij}, is the determinant of the matrix obtained from A by deleting the ith row and jth column. The **cofactor** of a_{ij} is the scalar C_{ij}, where $C_{ij} = (-1)^{i+j}M_{ij}$.

Now we can give the **Laplace expansion** for $\det(A)$. For any i between 1 and n we have

$$\det(A) = \sum_{j=1}^{n} a_{ij}C_{ij} = \sum_{j=1}^{n} a_{ji}C_{ji}.$$

If we use elementary row operations to reduce the matrix to one where the first row (say) has one non-zero entry — keeping track of the change this makes to the determinant! — then the problem of determining the determinant of the matrix can be reduced (using the Laplace transformation) to that for a smaller (and hence simpler) matrix.

We can also use the cofactors to write down the inverse of an invertible matrix A, in the following manner. The **adjoint** of A (written $\text{adj}\,(A)$) is the transpose of the matrix formed by replacing each a_{ij} with its cofactor. That is, it is the matrix B where $b_{ij} = C_{ji}$. Last year you should have seen that, when A is invertible, we have

$$A^{-1} = \frac{1}{\det(A)}\text{adj}\,(A).$$

The above formula gives us a method for calculating the inverse of a matrix, but it can be quite lengthy to implement. It is often simpler to use elementary row operations to reduce the augmented matrix $(A|I_n)$ to the form $(I_n|B)$; if this is possible then A is invertible, and its inverse is B.