mpc2w6.tex

Week 6. 16.11.2011

Identity Matrix.

Recall the *Kronecker delta* δ_{ij}:

$$\delta_{ij} := 1 \text{ if } i = j, \quad 0 \text{ otherwise}$$

(Leopold KRONECKER (1823-1891); posth. book 1903).
The $n \times n$ matrix I, or I_n, with (i, j) element δ_{ij},

$$I = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & & 1 \end{pmatrix}$$

is called the *identity matrix.*
Matrix Products.

Suppose

$$Ax = d$$

as above. Now make a change of variable $x \mapsto u$, where

$$\begin{aligned} x_1 &= b_{11}u_1 + \ldots + b_{1n}u_n, \\ \vdots &= \vdots \\ x_n &= b_{1n}u_1 + \ldots + b_{nn}u_n, \end{aligned}$$

or in matrix notation

$$x = Bu,$$

where B is the square matrix $B = (b_{ij})$. Then

$$x_j = \sum_k b_{jk}u_k,$$

so the above equation(s)

$$\sum_j a_{ij}x_j = d_i$$

become

$$\sum_j a_{ij} \sum_k b_{jk}u_k = d_i \qquad (i = 1, \ldots, n).$$

Write

$$c_{ik} := \sum_j a_{ij} b_{jk} \qquad (i, k = 1, \ldots, n). \tag{$*$}$$

Then this becomes

$$\sum_k c_{ik} u_k = d_i \qquad (i = 1, \ldots, n).$$

In matrix notation, this is

$$Cu = d,$$

with u, d n-vectors (columns), and $C = (c_{ij})_{i,j=1}^n$ the square matrix of order n given by $(*)$. We call C the *matrix product* of A and B, in that order:

$$C =: AB.$$

So we can now combine

$$Ax = d$$

and

$$x = BU$$

as

$$Ax = A(Bu) = (AB)u = Cu = d,$$

as above. So:

If $A = (a_{ij})$, $B = (b_{ij})$, $C = AB = (c_{ij})$, where $c_{ij} = \sum_k a_{ik} b_{kj}$.

Note. 1. Learn the pattern of suffices here. It does not matter which of i, j, k we choose. What matters is the *linking via consecutive suffices*, as above.
2. In the Einstein summation convention, this becomes simply

$$c_{ij} = a_{ik} b_{kj}$$

(summation over k understood). Make your own personal choice about when to use this – always, never, or when it seems convenient.
3. The matrices A, B need not be square, but they must be *conformable*: if A is $m \times n$, B is $n \times p$, then $C = AB$ is $m \times p$. Matrix multiplication is *row by column*: the number of *rows* of the first factor must be the same as the number of *columns* of the second factor. (In Ch. VI we will deal with *dot products*, or *inner products*, of vectors. This is an example; such products

are only defined for vectors of the same length.)
4. If A is $m \times n$, B is $n \times p$, C is $p \times q$, ABC is $m \times q$, defined as

$$ABC := A(BC), \text{ or } (AB)C.$$

These two are the same:

$$((AB)C)_{ij} = \sum_k (AB)_{ik} C_{kj} = \sum_k \sum_\ell a_{i\ell} b_{\ell k} c_{kj} = \sum_\ell a_{i\ell} \sum_k (BC)_{\ell j} = (A(BC))_{ij}.$$

That is, matrix multiplication is *associative*.

So we do not need brackets when writing matrix products. So we should not *use* brackets when writing matrix products.

We should use brackets when needed to avoid ambiguity (a sin in Mathematics), but not otherwise (as we should always use the lightest notation that does the job). Similarly with mathematical notation generally!
5. But matrix multiplication is not *commutative*. Even if AB and BA are both defined (both products are conformable – this can happen only if A and B are both square and of the same size),

$$AB \neq BA$$

in general. If one has $AB = BA$, one says that A and B *commute*.
6. This non-commutativity is essential in Quantum Mechanics, e.g. in the *Heisenberg commutation relations*

$$[p, q] := pq - qp = i\hbar$$

(Werner HEISENBERG (1901-1970), in 1925. (Heisenberg's formulation of Quantum Mechanics is called Matrix Mechanics. When Heisenberg derived it, he didn't know what a matrix was!)
7. We can display a matrix A as a row of its columns, $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ (or as a column of its rows). The kth column of the matrix product $C = AB$ is then

$$\mathbf{c}_k = b_{1k}\mathbf{a}_1 + \ldots + b_{nk}\mathbf{a}_n.$$

For, the ith element of the kth column of C is

$$c_{ik} = \sum_j a_{ij} b_{jk} = \sum_j b_{jk} [\mathbf{a}_j]_i = [\sum_j b_{jk} \mathbf{a}_j]_i.$$

This is the ith element of the above vector equation, on both sides. //
8. Determinants can be traced back to Leibniz (1684, unpublished in his

lifetime), Cramer (below) and others; the term first appears in Gauss' thesis *Disquisitiones arithmeticae* in 1801. Although matrices logically precede determinants, they were developed after them. The term is due to J. J. SYLVESTER (1814-1897) in 1850; the theory largely stems from a paper of Arthur CAYLEY (1821-1895) in 1858 (this contains the Cayley-Hamilton Theorem [Week 7], following work by Hamilton in 1853).

Multiplication Theorem.
Th.

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

Proof. If $C := AB$, by (7) above,

$$detC = detAB = det[b_{11}\mathbf{a}_1 + \ldots + b_{n1}\mathbf{a}_n, \ldots, b_{1n}\mathbf{a}_1 + \ldots + b_{nn}\mathbf{a}_n].$$

Expand the RHS by the first column. We get a sum of the form

$$\sum_{j_1} b_{j_1,1} det[\ldots].$$

Expand each det here by the second column. We get a double sum, of the form

$$\sum_{j_1,j_2} b_{j_1,1} b_{j_2,1} det[\ldots],$$

and so on, finally getting

$$\sum_{j_1,\ldots,j_n} b_{j_1,1} \ldots b_{j_n,1} det[\ldots].$$

Each matrix whose det we are taking here is a row of columns of A. Each such det with two columns the *same* vanishes. So we can reduce the 'big' sum (n^n terms) to a smaller sum with all columns *different* ($n!$ terms). Then we have a *permutation* of the columns, σ say, giving

$$detC = \sum_{\sigma} b_{\sigma(1),1} \ldots b_{\sigma(n),n} det[\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(n)}].$$

Putting the columns here in their natural order,

$$detC = \sum_{\sigma} b_{\sigma(1),1} \ldots b_{\sigma(n),n}.(-1)^{sgn(\sigma)} det[\mathbf{a}_1, \ldots, \mathbf{a}_n].$$

The determinant here is $detA$, so we can take it out. This leaves $detB$, so

$$detC = det(AB) = detA.detB. \qquad //$$

Adjoint.

If $A = (a_{ij})$ is $n \times n$:
the *minor* $|A_{ij}|$ of a_{ij} is the determinant of the matrix A_{ij} obtained by deleting the row and column containing a_{ij};
the *cofactor* of a_{ij} is the *signed minor* $\pm|A_{ij}|$ (the sign depends on whether the position (i,j) gets a + or a - sign in the expansion of the determinant – think of + as white and - as black, on a chessboard with white in the top left-hand corner);
the *adjoint* $adjA$ of A is the $n \times n$ matrix with (i,j) entry

$$(adjA)_{ij} := \pm|A_{ji}| :$$

the adjoint is the transposed matrix of cofactors.

Th.

$$A.adjA = adjA.A = |A|.I.$$

Proof.

$$(A.adjA)_{ij} = \sum_k a_{ik}(adjA)_{kj} = \sum_k a_{ik}. \pm |A_{jk}|.$$

Recall that $detA = \sum_k a_{ik}. \pm |A_{ik}|$, expanding $detA$ by the ith row, So this is the determinant, expanded by the ith row, of the matrix obtained from A by replacing its ith row by its jth row. This determinant is 0 if $i \neq j$ (det of a matrix with two identical rows vanishes), but if $i = j$ it is $detA$. Combining,

$$(A.adjA)_{ij} = |A|.\delta_{ij} = |A|.I_{ij},$$

giving

$$A.adjA = |A|.I,$$

and similarly

$$(adjA).A = |A|.I. \qquad //$$

Inverse Matrix. Call a square matrix A *singular* if $|A| = 0$, *non-singular* if $|A| \neq 0$. The *inverse matrix* A^{-1} of A is undefined if A is singular; for A non-singular,

$$A^{-1} := \frac{adjA}{|A|}.$$

So by above,
$$AA^{-1} = A^{-1}A = I.$$
Solution of Linear Equations.

For A square,
$$Ax = b$$
has a unique solution x iff A is *non-singular*,
$$|A| \neq 0.$$
Then the unique solution is
$$x = A^{-1}b.$$
For, just pre-multiply by A^{-1}: as $A^{-1}A = I$,
$$A^{-1}Ax = Ix = x = A^{-1}b. \qquad //$$

Note. If A is singular, $|A| = 0$, A^{-1} does not exist, and there are either *no* solutions or *infinitely many*, depending on the *rank* (order of the largest non-zero minor) of the *augmented matrix* (A, b). We omit further detail.

Cramer's Rule (G. CRAMER (1704-1752) in 1750).
Th. If $|A| = 0$, the solution x to $Ax = b$ is given by
$$x_i = \Delta_i/\Delta,$$
where $\Delta = |A|$ and Δ_i is the det of the matrix A_i obtained from A by replacing its ith column by the RHS b.
Proof.
$$(A^{-1}b)_i = \sum_j (A^{-1})_{ij} b_j = \sum_j \pm |A_{ji}| b_j / |A|$$
$$= \sum_j \pm b_j |A_{ji}| / |A| = |A_i|/|A| = \Delta_i/\Delta,$$
expanding Δ_i by its ith column. //

Gaussian Elimination.

Although Cramer's Rule is theoretically neat, it does not provide an efficient way of actually calculating the solution numerically. For this, we need a procedure such as Gaussian elimination, below.

Given the equations we encountered earlier, $Ax = b$ or in scalar notation

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ \vdots &= \vdots \\ a_{1n}x_1 + \ldots + a_{nn}x_n &= b_n. \end{aligned}$$

To solve this, we find a coefficient of x_1 (say) which is non-zero (if there is none, x_1 is missing, and we can eliminate it and one equation). We can take this as a_{11} – if not, re-order the equations. Divide the first equation by $a_{11} \neq 0$, to make the coefficient of x_1 1.

Note. If a_{11} is small (near 0), dividing by it is numerically unstable (small rounding errors in a_{11} would be magnified – think of relative v. absolute error here!). For maximum numerical stability, choose as a_{11} the coefficient (a_{1i}, say) of largest modulus, and bring this to the $(1, 1)$ position by re-ordering the equations.

We now have equations of the form

$$\begin{aligned} x_1 + \ldots + a_{1n}x_n &= c_1, \\ (a_{21}/a_{11})x_1 + \ldots &= c_2, \\ \vdots &= \vdots \\ (a_{1n}/a_{11})x_1 + \ldots &= c_n. \end{aligned}$$

Subtract the appropriate multiple (a_{i1}/a_{11}) of the first equation from the ith equation. This *eliminates* x_1 from the ith equation for $i > 1$, giving a set of equations whose matrix has only 0s below the diagonal in the first column.

Now leave the first equation (n variables), and focus on the remaining $n - 1$ equations (in $n - 1$ variables). Treat this new set as we treated the old set – and *repeat* the operation, $n - 1$ times. As at each step we obtain 0s below the diagonal in the current first row, we end up with a set of equations whose matrix is *lower triangular* – has only 0s below the diagonal. For such a lower triangular system, we can solve easily, by *back-substitution.* The last equation has only one unknown; solve for it. Then back-substitute in the next one up, which then has only one unknown; solve for it. Repeating this, we solve the entire system.

This method of *Gaussian elimination* was used by C. F. GAUSS (1777-1855) (c. 1805, calculating the orbit of Pallas, and 1809, of Ceres).

Numerical Linear Algebra.

This is a subject in its own right. Also, some computer packages/programming languages are specifically geared to matrix computation, e.g. MATLAB.

Similarity.

Imagine a curve – ellipse, say – or surface – ellipsoid, say. It is still an ellipse/ellipsoid if we change coordinates. We now study such changes of coordinates.

Take a linear system of equations

$$y = Ax,$$

and a new coordinate system

$$x = Px'$$

(x' to save using a new letter – no differentiation involved!), where P is non-singular (so we can change back via $x' = P^{-1}x$). The new system is $y' = A'x'$, or

$$A'x' = y' = P^{-1}y = P^{-1}Ax = P^{-1}APx'.$$

So

$$A' = P^{-1}AP.$$

Such matrices A, A' are said to be *similar.*

Diagonalization.

Suppose we want to change A by a similarity transformation to a diagonal matrix D (as in reducing an ellipse to standard form, say). Then as above $D = P^{-1}AP$, or

$$AP = PD.$$

If

$$P = [p_1, \ldots, p_n] \quad \text{row of columns}, \quad D = diag(d_1, \ldots, d_n),$$

this is

$$A[p_1, \ldots, p_n] = [p_1, \ldots, p_n] \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = [d_1 p_1, \ldots, d_n p_n],$$

or

$$Ap_i = d_i p_i \qquad (i = 1, \ldots, n).$$

This has the form of an *eigenvalue problem*; see Week 7.