Vectors and matrices: matrices (Version 2)

This is a very brief summary of my lecture notes.

Matrices and linear equations

A *matrix* is an m-by-n array of numbers

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}.$$

A system of linear equations

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{$*$}$$

gives rise to a *coefficient matrix*

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

and an *augmented matrix*

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

An *elementary row operation (ERO)* on a matrix is one of the following operations:

- multiply a row by a nonzero scalar,

- swap two rows,
- add a scalar multiple of one row to another (**different**) row.

The effects of an ERO can be reversed by applying an ERO.

Applying an ERO to the augmented matrix of a system $(*)$ of equations gives rise to the augmented row of an **equivalent** system $(*')$ of equations. (The fact that EROs are reversible is crucial for this.) Hence we can solve $(*)$ by performing EROs on its augmented matrix, transforming it into a convenient form.

Each matrix can be transformed into *echelon form.* This is a matrix of the form

$$\begin{pmatrix}
0 & \cdots & 0 & \dagger & * & \cdots & * & * & * & \cdots & * & * & * & \cdots & * & * & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \dagger & * & \cdots & * & * & * & \cdots & * & * & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \dagger & * & \cdots & * & * & * & \cdots & * \\
 & \ddots & & & & \ddots & & & & \ddots & & & & \ddots & & & & \ddots & \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \dagger & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
 & \ddots & & & & \ddots & & & & \ddots & & & & \ddots & & & & \ddots & \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0
\end{pmatrix}.$$

Here the entries $\dagger$ are nonzero and are called the *pivots.* The entries $*$ can be any numbers. The pivots are all in distinct columns. All the rows below the last pivot (if there are any) are zero. A matrix is in *reduced echelon form* if it is in echelon form, all its pivots are equal to 1 and all the entries in the column above each pivot are 0. While a matrix can be transformed into many different echelon form matrices, it may be transformed into precisely one reduced echelon form matrix.

If we have a system of equations $(*)$ whose augmented matrix is in echelon form we can easily find its set of solutions. First of all, if the final pivot is in the last column, the relevant row corresponds to the equation

$$0x_1 + 0x_2 + \cdots + 0x_n = b_r$$

where $b_r \neq 0$. But as this equation asserts that $0 = b_r$ the system has no solutions.

If the final pivot is not in the last column, the system is soluble. The matrix will have r nonzero rows, each with one pivot. These pivots are in r columns corresponding to r variables, which we call *basic variables.* The other $n - r$ variables are the *free variables.* By working from the bottom up, we can express each basic variable in terms of the free variables. We thus get an $(n-r)$-parameter system of solutions—the *general solution*. Substituting values for the free variables gives a *particular* solution. The case where $n = r$ is of special interest. Here all variables are basic, and there is a unique solution to the system.

Matrix arithmetic

An matrix with one column is a *column vector* and a matrix with one row is a *row vector*. By tradition we use mainly column vectors. If

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

is an m-by-n matrix and

$$\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

is a column vector with n entries, we define

$$A\mathbf{v} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

This is an column vector with m entries. Thus we can multiply an m-by-n matrix by an n-entry column vector to yield an m-entry column vector.

We define matrix addition, subtraction and multiplication in order to make the identities

$$(A+B)\mathbf{v} = A\mathbf{v} + B\mathbf{v}, \qquad (A-B)\mathbf{v} = A\mathbf{v} - B\mathbf{v}, \qquad (AB)\mathbf{v} = A(B\mathbf{v})$$

hold as widely as possible. For addition and subtraction, the matrices A and B must have the same size, m-by-n, and then $A+B$ and $A-B$ are also m-by-n. They are easy to calculate: the (i,j)-entry of $A+B$ is $a_{ij}+b_{ij}$ where a_{ij} and b_{ij} are the (i,j)-entries of A and B respectively. Similarly $A-B$ has (i,j)-entry $a_{ij} - b_{ij}$. We can also define multiplication by a scalar: cA has (i,j)-entry ca_{ij}.

Multiplication is a lot trickier! A product AB exists only if the number of columns of A equals the number of rows of B, that is, if A is m-by-n and

B is n-by-p for some m, n and p. In this case, AB is an m-by-p matrix whose (i, j)-entry is

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

It may help to remember that this is the entry in row i and column j of AB and it depends on row i of A and column j of B; indeed it is the product of this row and column, considered as vectors.

The following identities involving matrix multiplication hold:

$$(A + B)C = AC + BC, \qquad A(B + C) = AB + AC, \qquad A(BC) = (AB)C.$$

But it is **false** that $AB = BA$ in general. Matrix multiplication is *noncommutative.* All of the following cases occur:

- AB exists but BA doesn't,
- AB and BA both exist but they have different sizes,
- AB and BA both exist and have the same size, but $AB \neq BA$.

For each n there is an n-by-n *identity matrix* I_n. This satisfies $I_nA = A$ for all n-row matrices A and $BI_n = B$ for all n-column matrices B. Then

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

In detail, the entries on the main diagonal equal 1 and all others equal 0. Usually we write I for any I_n.

A matrix B is the *inverse* of a matrix A if $AB = I$ and $BA = I$. A matrix is *invertible* or *nonsingular* if it has an inverse. Only square (n-by-n) matrices can be invertible, and an invertible matrix has precisely one inverse. For if B and C are inverses of A then

$$B = BI = B(AC) = (BA)C = IC = C.$$

It's a theorem that if A is square and $AB = I$ then $BA = I$ also. But one can have nonsquare matrices A and B such that AB is an identity matrix; however then BA won't be an identity. If B is the inverse of A, then we write $B = A^{-1}$.

The product AB of two nonsingular matrices A and B is nonsingular. For

$$(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I$$

and

$$(B^{-1}A^{-1})(AB) = B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I$$

so that AB has the inverse $B^{-1}A^{-1}$.

To compute an inverse of an n-by-n matrix A, make a new n-by-$2n$ matrix $M = (A\ I)$ by putting M next to the identity, then convert M into reduced echelon form N by elementary row operations. If $N = (I\ B)$ then $B = A^{-1}$; otherwise A is not invertible.

The *transpose* of an m-by-n matrix A is the n-by-m matrix A^t obtained from A by interchanging rows with columns. That is

$$\text{if} \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{then} \quad A^t = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

Transposition preserves addition and subtraction: $(A + B)^t = A^t + B^t$ and $(A - B)^t = A^t - B^t$, but reverses order of multiplication: $(AB)^t = B^tA^t$. A *symmetric matrix* is a (square) matrix A satisfying $A^t = A$; a *skew-symmetric matrix* is a (square) matrix B satisfying $B^t = -B$.

An *elementary* matrix is a matrix E obtained by applying an elementary row operation to an identity matrix. Each elementary matrix is invertible and its inverse is also an elementary matrix. If A is an m-by-n matrix, then applying an elementary row operation to A yields the matrix EA where E is the elementary matrix obtained by performing the same elementary row operation on I_m. As a consequence, there is a sequence $E_1, \ldots, E_k$ of elementary matrices with $E_kE_{k-1} \cdots E_2E_1A = B$ in (reduced) echelon form. If A is invertible, then we can take $B = I$ and get that $A = E_1^{-1}E_2^{-1} \cdots E_{k-1}^{-1}E_k^{-1}$ is a product of elementary matrices.

Geometry of matrix transformations

For simplicity we work in the plane. If A is a 2-by-2 matrix and $P = (x, y)$ is a point in the plane define $P' = (x', y')$ where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}.$$

The mapping $P \mapsto P'$ is a transformation of the plane to itself which preserves the origin, and if A is nonsingular, takes lines to lines and parallelograms to parallelograms. Some examples:

- $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ for $a > 0$ is a magnification with factor a,
- $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$ for $a > 0$ is a stretch with factor a in the x-direction,
- $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is reflection in the x-axis,
- $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is reflection in the line $y = x$,
- $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is rotation through angle θ anticlockwise about the origin.

Determinants

Each square matrix has A a *determinant*. This is a number associated to A in a certain way and is denoted by $|A|$ or $\det(A)$. There are several ways of defining determinants. In this course we give a recursive method, defining n-by-n determinants in terms of $(n-1)$-by-$(n-1)$ determinants.

The formula for a 1-by-1 determinant is trivial:

$$|a| = a.$$

The formula for a 2-by-2 determinant is well-worth remembering:

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

It's sometimes handy, but not essential to remember the formula for a 3-by-3 determinant:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & j \end{vmatrix} = aej + bfg + cdh - afh - bdj - ceg.$$

The formula for an n-by-n determinant has $n!$ terms; attempting to memorize larger determinant formulae is completely bonkers!

To define the determinant in general let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

be a typical n-by-n matrix. For numbers j and k between 1 and n define A_{jk} to be the matrix obtained from A by deleting row j and column k. Then A_{jk} is an $(n-1)$-by-$(n-1)$ matrix. For instance, when $n = 4$ then

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \qquad \text{and} \qquad A_{31} = \begin{pmatrix} a_{12} & a_{13} & a_{14} \\ a_{22} & a_{23} & a_{24} \\ a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Now we define

$$|A| = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| - \cdots \pm a_{1n}|A_{1n}| = \sum_{k=1}^{n} (-1)^{k+1} a_{1k}|A_{1k}|.$$

We check that this definition yields the formulae given above for small matrices.

A consequence of this definition is that the determinant of an upper triangular matrix is the product of its diagonal elements. An *upper triangular matrix* is a square matrix with all entries below the main diagonal equal to zero. Thus

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}a_{33}\cdots a_{nn}.$$

For large matrices this definition is inefficient for calculation. Elementary row operations have predictable effects on determinants. Suppose that matrix A is transformed to matrix B via an elementary row operation:

- if B arises by multiplying a row of A by a scalar t then $|B| = t|A|$, equivalently, $|A| = t^{-1}|B|$;
- if B arises by swapping two rows of A then $|B| = -|A|$;
- if B arises by adding a scalar multiple of a row of A to another row of A then $|B| = |A|$.

A square matrix in echelon form is upper triangular. Hence one may calculate determinants by reducing a matrix to echelon form using row operations.

We now list some facts about determinants:

- a matrix with an all-zero row or all-zero column has zero determinant;

- determinants are invariant under transposition: $|A| = |A^t|$;
- the determinant is multiplicative: $|AB| = |A||B|$;
- as $|I| = 1$ it follows that when $|A|$ is nonsingular, $|A| \neq 0$ and $|A^{-1}| = |A|^{-1}$ (since $|A||A^{-1}| = |AA^{-1}| = |I| = 1$);
- by contrast, if A is singular, $|A| = 0$, so $|A| = 0$ is a necessary and sufficient for A to be singular.

The formula defining $|A|$ in terms of $a_{11}, \dots, a_{1n}$ and $|A_{11}|, \dots, |A_{1n}|$ is known as expansion of the determinant along the first row. We can expand the determinant along other rows, for instance the second row:

$$|A| = -a_{21}|A_{21}| + a_{22}|A_{22}| - a_{23}|A_{23}| + \cdots \pm a_{2n}|A_{2n}| = \sum_{k=1}^{n} (-1)^{k+2} a_{2k}|A_{2k}|$$

etc. The formula for expansion along row j is

$$|A| = \sum_{k=1}^{n} (-1)^{j+k} a_{jk}|A_{jk}|.$$

Similarly we can expand along columns; expanding down column i gives

$$|A| = \sum_{k=1}^{n} (-1)^{i+k} a_{ki}|A_{ki}|.$$

A consequence

$$A \operatorname{adj} A = |A| I$$

where $\operatorname{adj} A$ is the matrix whose entry in row i and column j is $(-1)^{i+j}|A_{ji}|$. Note that $|A_{ji}|$ is the determinant of the matrix obtained by deleting **column** i and **row** j from A. Also the signs $(-1)^{i+j}$ form a chessboard pattern. Hence if A is non-singular, then

$$A^{-1} = \frac{1}{|A|} \operatorname{adj} A.$$

This formula for the inverse of a matrix is practical for 2-by-2 and sometimes 3-by-3 matrices but rarely for larges matrices.

Eigenvectors and eigenvalues

Let A be a square matrix. A vector $\mathbf{v}$ is an *eigenvector* of A if $\mathbf{v} \neq 0$ and $A\mathbf{v} = t\mathbf{v}$ for some scalar t. The number t is called the *eigenvalue* associated to the eigenvector $\mathbf{v}$. We can rewrite the equation $A\mathbf{v} = t\mathbf{v}$ as

$$(tI - A)\mathbf{v} = 0. \qquad (*)$$

For t to be an eigenvalue of A, $(*)$ must have a non-zero vector $\mathbf{v}$ as a solution. This happens if and only if the matrix $tI - A$ is singular, that is, it has determinant zero. Hence t is an eigenvalue of A if and only if

$$|tI - A| = 0. \tag{$\dagger$}$$

If A is n-by-n, $(\dagger)$ is a degree n algebraic equation. It is called the *characteristic equation* of A. To find the eigenvalues and eigenvectors of A we first solve the characteristic equation $(\dagger)$ to find the eigenvalues t; then for each eigenvalue we solve $(*)$ to find its associated eigenvectors. In each case we will get a family of solutions depending on one or more parameters.

If the characteristic equation of A has n distinct roots $t_1, \ldots, t_n$ we can build an n by n matrix V with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then AV has columns $t_1\mathbf{v}_1, \ldots, t_n\mathbf{v}_n$. Hence

$$AV = VD \qquad \text{where} \qquad D = \begin{pmatrix} t_1 & 0 & 0 & \cdots & 0 \\ 0 & t_2 & 0 & \cdots & 0 \\ 0 & 0 & t_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_n \end{pmatrix}$$

is a *diagonal* matrix (a matrix whose nonzero entries are all on the main diagonal). It is a fact that when the characteristic equation has distinct roots then this matrix V is nonsingular, and so

$$D = V^{-1}AV.$$

We say that A is *diagonalizable* if there is a non-singular matrix V with $V^{-1}AV$ diagonal. All n-by-n matrices with n distinct eigenvalues are diagonalizable; matrices with repeated eigenvalues may or may not be diagonalizable.

If we can diagonalize A it is easy to compute A^k for all k. If

$$D = \begin{pmatrix} t_1 & 0 & 0 & \cdots & 0 \\ 0 & t_2 & 0 & \cdots & 0 \\ 0 & 0 & t_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_n \end{pmatrix} \qquad \text{then} \qquad D^k = \begin{pmatrix} t_1^k & 0 & 0 & \cdots & 0 \\ 0 & t_2^k & 0 & \cdots & 0 \\ 0 & 0 & t_3^k & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_n^k \end{pmatrix}$$

and if $D = V^{-1}AV$ then $A = VDV^{-1}$ and

$$A^k = VDV^{-1}VDV^{-1} \cdots VDV^{-1} = VDIDI \cdots IDV^{-1} = VD^kV^{-1}.$$

Matrices A which have repeated eigenvalues may not be diagonalizable. If t is an eigenvalue of A its *algebraic multiplicity* $\mu_a(t)$ is its multiplicity as a root of the characteristic equation of A. Its *geometric multiplicity* $\mu_g(t)$ is the number of parameters needed to express all eigenvectors with eigenvalue t, equivalently, the number of nonzero rows in the echelon form for $tI - A$. It is a fact that $1 \leq \mu_g(t) \leq \mu_a(t)$ for all eigenvalues t. Also A is diagonalizable if and only if $\mu_g(t) = \mu_a(t)$ for all eigenvalues t. A simple example is the matrix

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

This has the sole eigenvalue 1 with $\mu_a(1) = 2$ but $\mu_g(1) = 1$; A is not diagonalizable.

If A is diagonalizable, with repeated eigenvalues, we can construct the matrix V diagonalizing A as follows. For each eigenvalue t let $\mu = \mu_a(t) = \mu_g(t)$. Then the equation $(tI - A)\mathbf{v} = 0$ has general solution $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_\mu\mathbf{v}_\mu$. Insert columns $\mathbf{v}_1 \ldots, \mathbf{v}_\mu$ into the matrix V, and do this for all eigenvalues t.

RJC 11/1/2005