a quick rundown on determinants (for MA 340 and 443)

Determinants are often brought into courses such as this quite unnecessarily. But when they are useful, they are remarkably so. The use of determinants is a bit bewildering to the beginner, particularly if confronted with the classical definition as a sum of signed products of matrix entries.

I find it more intuitive to follow Weierstrass and begin with a few important properties of the determinant, from which all else follows, including that classical definition (which is practically useless anyway).

(As to the many determinant identities available, in the end I have always managed with just one nontrivial one, viz. *Sylvester's determinant identity*, and this is nothing but Gauss elimination; see the end of these notes. The only other one I have often used is the *Cauchy-Binet formula*.)

The determinant is a map,

$$\det : \mathbb{F}^{n\times n} \to \mathbb{F} : A \mapsto \det A,$$

with various properties. The first one in the following list is perhaps the most important one.

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

(ii) $\det(I) = 1$

Consequently, for any invertible A,

$$1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1}).$$

Hence,

(iii) *If A is invertible, then* $\det A \neq 0$ *and,* $\det(A^{-1}) = 1/\det(A)$.

While the determinant is defined as a map on matrices, it is very useful to think of $\det(A) = \det[\mathbf{a}_1, \ldots, \mathbf{a}_n]$ as a function of the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of A. The next two properties are in those terms:

(iv) $\mathbf{x} \mapsto \det[\ldots, \mathbf{a}_{j-1}, \mathbf{x}, \mathbf{a}_{j+1}, \ldots]$ *is linear*, i.e., for any n-vectors $\mathbf{x}$ and $\mathbf{y}$ and any scalar α (and arbitrary n-vectors $\mathbf{a}_i$),

$$\det[\ldots, \mathbf{a}_{j-1}, \mathbf{x} + \alpha\mathbf{y}, \mathbf{a}_{j+1}, \ldots] = \det[\ldots, \mathbf{a}_{j-1}, \mathbf{x}, \mathbf{a}_{j+1}, \ldots] + \alpha\det[\ldots, \mathbf{a}_{j-1}, \mathbf{y}, \mathbf{a}_{j+1}, \ldots].$$

(v) *The determinant is an* **alternating** *form*, i.e.,

$$\det[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots] = -\det[\ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots].$$

In words: Interchanging two columns changes the sign of the determinant (and nothing else).

It can be shown (see below) that (ii) + (iv) + (v) implies (i) (and anything else you may wish to prove about determinants). Here are some basic consequences first.

(vi) Since 0 is the only scalar α with the property that $\alpha = -\alpha$, it follows from (v) that $\det(A) = 0$ *if two columns of A are the same.*

(vii) *Adding a multiple of one column of A to another column of A doesn't change the determinant.*

Indeed, using first (iv) and then the consequence (vi) of (v), we compute

$$\det[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j + \alpha\mathbf{a}_i, \ldots] = \det[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots] + \alpha\det[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_i, \ldots] = \det[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots].$$

Here comes a very important use of (vii): Assume that $\mathbf{b} = A\mathbf{x}$ and consider $\det[\ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots]$. Since $\mathbf{b} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$, subtraction of x_i times column i from column j, i.e., subtraction of $x_i\mathbf{a}_i$ from $\mathbf{b}$ here, for each $i \neq j$ is, by (vii), guaranteed not to change the determinant, yet changes the jth column to $x_j\mathbf{a}_j$; then, pulling out that scalar factor x_j (permitted by (iv)), leaves us finally with $x_j \det A$. This proves

(viii) *If* $\mathbf{b} = A\mathbf{x}$, *then*

$$\det[\ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots] = x_j \det A.$$

Hence, if $\det A \neq 0$, then $\mathbf{b} = A\mathbf{x}$ implies

$$x_j = \det[\ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots]/\det(A), \qquad j = 1, \ldots, n.$$

This is **Cramer's rule**.

In particular, it follows that $x_j = 0$ for all j, in case $A\mathbf{x} = \mathbf{0}$ and $\det(A) \neq 0$. By the Invertible Matrix Theorem, this gives the converse to (iii), i.e.,

(ix) *If* $\det(A) \neq 0$, *then* A *is invertible.*

In old-fashioned mathematics, a matrix was called **singular** if its determinant is 0. So, (iii) and (ix) combined say that a matrix is nonsingular iff it is invertible.

The suggestion that one actually construct the solution to $A? = b$ by Cramer's rule is ridiculous under ordinary circumstances since, even for a linear system with just two unknowns, it is more efficient to use Gauss elimination. On the other hand, if the solution is to be constructed *symbolically* (in a symbol-manipulating system such as `Maple` or `Mathematica`), then Cramer's rule is preferred to Gauss elimination since it treats all unknowns equally. In particular, the number of operations needed to obtain a particular unknown is the same for all unknowns.

We have proved all these facts (except (i)) about determinants from certain postulates (namely (ii), (iv), (v)) without ever saying how to *compute* $\det(A)$. Now, it is the actual formulas for $\det(A)$ that have given determinants such a bad name. Here is the standard one, which (see below) can be derived from (ii), (iv), (v), in the process of proving (i):

(x) *If* $A = (a_{ij} : i, j = 1, \ldots, n)$, *then*

$$\det(A) = \sum_{\sigma \in \$_n} (-1)^\sigma \prod_{j=1}^{n} a_{\sigma(j),j}$$

Here, $\sigma \in \$_n$ is shorthand for: σ is a **permutation of the first** n **integers**, i.e.,

$$\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)),$$

where $\sigma(j) \in \{1, 2, \ldots, n\}$ for all j, and $\sigma(i) \neq \sigma(j)$ if $i \neq j$. In other words, σ is a 1-1 and onto map from $\{1, \ldots, n\}$ to $\{1, \ldots, n\}$. This is bad enough, but I still have to explain the mysterious $(-1)^\sigma$. This number is 1 or -1 depending on whether the **parity** of σ is even or odd. Now, this parity can be determined in at least two equivalent ways:

(a) keep making interchanges until you end up with the sequence $(1, 2, \ldots, n)$; the parity of the number of steps it took is the parity of σ (note the implied assertion that this parity will not depend on how you went about this, i.e., the number of steps taken may differ, but the parity never will; if it takes me an even number of steps, it will take you an even number of steps.)

(b) count the number of pairs that are out of order; its parity is the parity of σ.

Here is a simple example: $\sigma = (3, 1, 4, 2)$ has the pairs $(3, 1)$, $(3, 2)$, and $(4, 2)$ out of order, hence $(-1)^\sigma = -1$. Equivalently, the following sequence of 3 interchanges gets me from σ to $(1, 2, 3, 4)$:

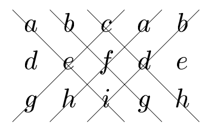
$$\begin{array}{c} (3,1,4,2) \\ (3,1,2,4) \\ (1,3,2,4) \\ (1,2,3,4) \end{array}$$

Therefore, again, $(-1)^\sigma = -1$.

Now, fortunately, we don't really ever have to use this stunning formula (x) in calculations, nor is it physically possible to use it for n much larger than 8 or 10. For $n = 1, 2, 3$, one can derive from it explicit rules for computing $\det(A)$:

$$\det[\,a\,] = a, \quad \det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc, \quad \det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - (ceg + afh + bdi);$$

the last one can be remembered easily by the following mnemonic:



For $n > 3$, this mnemonic *does not work*, and one would not usually make use of (x), but use instead (i) and the following immediate consequence of (x):

(xi) *The determinant of a triangular matrix equals the product of its diagonal entries.*

Indeed, when A is upper triangular, then $a_{ij} = 0$ whenever $i > j$. Now, if $\sigma(j) > j$ for some j, then the factor $a_{\sigma(j),j}$ in the corresponding summand $(-1)^\sigma \prod_{j=1}^n a_{\sigma(j),j}$ is zero. This means that the only possibly nonzero summands correspond to σ with $\sigma(j) \le j$ for all j, and there is only one permutation that manages that, the **identity permutation** $(1, 2, \ldots, n)$, and its parity is obviously even. Therefore, the formula in (x) gives $\det A = a_{11} \cdots a_{nn}$ in this case. – The proof for a lower triangular matrix is analogous; else, use (xiii) below.

Consequently, if $A = LU$ with L unit triangular and U upper triangular, then

$$\det A = \det U = u_{11} \cdots u_{nn}.$$

If, more generally, $A = PLU$, with P some permutation matrix, then

$$\det A = \det(P) u_{11} \cdots u_{nn},$$

i.e.,

(xii) $\det A$ *is the product of the pivots used in elimination, times* $(-1)^i$, *with* i *the number of row interchanges made.*

Since, by elimination, any $A \in \mathbb{F}^n$ can be factored as $A = PLU$, with P a permutation matrix, L unit lower triangular, and U upper triangular, (xii) provides the standard way to compute determinants.

Note that, then, $A^T = U^T L^T P^T$, with U^T lower triangular, L^T unit upper triangular, and P^T the inverse of P, hence

(xiii) $\det A^T = \det A$.

This can also be proved directly from (x). Note that this converts all our statements about the determinant in terms of *columns* to the corresponding statements in terms of *rows*.

(xiv) "expansion by minors":

Since, by (iv), the determinant is slotwise linear, and $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$, we obtain

$$(1) \qquad \det[\ldots, \mathbf{a}_{j-1}, \mathbf{x}, \mathbf{a}_{j+1}, \ldots] = x_1 C_{1j} + x_2 C_{2j} + \cdots + x_n C_{nj},$$

with

$$C_{ij} := \det[\ldots, \mathbf{a}_{j-1}, \mathbf{e}_i, \mathbf{a}_{j+1}, \ldots]$$

the socalled **cofactor** of a_{ij}. With the choice $\mathbf{x} = \mathbf{a}_k$, this implies

$$a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk} = \det[\ldots, \mathbf{a}_{j-1}, \mathbf{a}_k, \mathbf{a}_{j+1}, \ldots] = \begin{cases} \det A, & k = j; \\ 0, & k \ne j. \end{cases}$$

The case $k = j$ gives the **expansion by minors** for $\det A$ (and justifies the name 'cofactor' for C_{ij}). The case $k \ne j$ is justified by (vi). In other words, with

$$\mathrm{adj} A := \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

the socalled **adjugate** of A (note that the subscripts appear reversed), we have

$$\mathrm{adj}(A)\, A = (\det A)\, I.$$

For an *invertible* A, this implies that

$$A^{-1} = (\mathrm{adj}A)/\det A.$$

The expansion by minors is useful since, as follows from (x), the cofactor C_{ij} equals $(-1)^{i+j}$ times the determinant of the matrix $A(\mathbf{n}\backslash i|\mathbf{n}\backslash j)$ obtained from A by removing row i and column j, i.e.,

$$C_{ij} = (-1)^{i+j} \det \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots \\ \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

and this is a determinant of order $n-1$, and so, if $n-1>1$, can itself be expanded along some column (or row).

(xv) $\det A$ *is the* n*-dimensional (signed) volume of the parallelepiped*

$$\{A\mathbf{x} : 0 \le x_i \le 1, \text{all } i\}$$

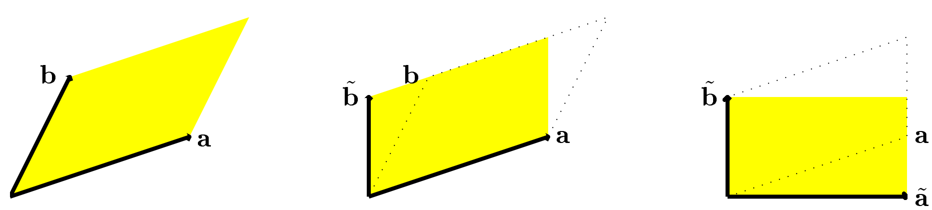
spanned by the columns of A.

For $n > 3$, this is a *definition*, while, for $n \le 3$, one works it out (see the book and/or else below). This is a very useful *geometric* way of thinking about determinants. Also, it has made determinants indispensable in the *definition* of multivariate integration and the handling therein of changes of variable.

Since $\det(AB) = \det(A)\det(B)$, it follows that *the linear transformation* $T : \mathbb{F}^n \to \mathbb{F}^n : \mathbf{x} \mapsto A\mathbf{x}$ *changes volumes by a factor of* $\det(A)$, meaning that, for any set M in the domain of T,

$$\mathrm{vol}_n(T(M)) = \det(A)\,\mathrm{vol}_n(M).$$

As an example, consider $\det[\mathbf{a}, \mathbf{b}]$, with $\mathbf{a}$, $\mathbf{b}$ vectors in the plane linearly independent, and assume, wlog, that $a_1 \ne 0$. By (iv), $\det[\mathbf{a}, \mathbf{b}] = \det[\mathbf{a}, \tilde{\mathbf{b}}]$, with $\tilde{\mathbf{b}} := \mathbf{b} - (b_1/a_1)\mathbf{a}$ having its first component equal to zero, and so, again by (iv), $\det[\mathbf{a}, \mathbf{b}] = \det[\tilde{\mathbf{a}}, \tilde{\mathbf{b}}]$, with $\tilde{\mathbf{a}} := \mathbf{a} - (a_2/\tilde{b}_2)\tilde{\mathbf{b}}$ having its second component equal to zero. Therefore, $\det[\mathbf{a}, \mathbf{b}] = \tilde{a}_1\tilde{b}_2 = \pm\|\tilde{\mathbf{a}}\|\|\tilde{\mathbf{b}}\|$ equals $\pm$ the area of the rectangle spanned by $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. However, following the derivation of $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ graphically, we see, by matching congruent triangles, that the rectangle spanned by $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ has the same area as the parallelepiped spanned by $\mathbf{a}$ and $\tilde{\mathbf{b}}$, and, therefore, as the paralellepiped spanned by $\mathbf{a}$ and $\mathbf{b}$. Thus, up to sign, $\det[\mathbf{a}, \mathbf{b}]$ is the area of the parallelepiped spanned by $\mathbf{a}$ and $\mathbf{b}$.



Here, finally, for the record, is a *proof* that (ii) + (iv) + (v) implies (i), hence everything else we have been deriving so far. Let A and B be arbitrary matrices (of order n). Then the linearity (iv) implies that

$$\det(BA) = \det[B\mathbf{a}_1, B\mathbf{a}_2, \ldots, B\mathbf{a}_n] = \det[\ldots, \sum_i \mathbf{b}_i a_{ij}, \ldots] = \sum_{\sigma\in\{1,\ldots,n\}^n} \det[\mathbf{b}_{\sigma(1)}, \ldots, \mathbf{b}_{\sigma(n)}] \prod_j a_{\sigma(j),j}.$$

By the consequence (vi) of the alternation property, most of these summands are zero. Only those determinants $\det[\mathbf{b}_{\sigma(1)}, \ldots, \mathbf{b}_{\sigma(n)}]$ for which all the entries of σ are different are *not* automatically zero. But that are exactly all the $\sigma \in \$_n$, i.e., the permutations of the first n integers. Further, for such σ,

$$\det[\mathbf{b}_{\sigma(1)}, \ldots, \mathbf{b}_{\sigma(n)}] = (-)^\sigma \det(B)$$

by the alternation property, with $(-)^\sigma = \pm 1$ depending on whether it takes an even or an odd number of interchanges to change σ into a strictly increasing sequence. (We discussed this earlier; the only tricky part remaining here is an argument that shows the parity of such number of needed interchanges to be independent of how one goes about making the interchanges. The clue to the proof is the simple observation that any one interchange is bound to change the number of sequence entries out of order by an *odd* amount.) Thus

$$\det(BA) = \det(B) \sum_{\sigma\in\$_n} (-)^\sigma \prod_j a_{\sigma(j),j}.$$

Since $IA = A$ while, by the defining property (ii), $\det(I) = 1$, the formula (x) follows and, with that, $\det(BA) = \det(B)\det(A)$ for arbitrary B and A. On the other hand, starting with the formula in (x) as a definition, one readily verifies that det so defined satisfies the three properties (ii) ($\det(I) = 1$), (iv) (multilinear), and (v) (alternating) claimed for it. In other words, there actually is such a function (necessarily given by (x)).

Here, for the record (but not discussed in class) is a proof and statement of Sylvester's Determinant Identity. For it, the following notation will be useful: If $\mathbf{i} = (i_1,\ldots,i_r)$ and $\mathbf{j} = (j_1,\ldots,j_s)$ are suitable integer sequences, then $A(\mathbf{i}|\mathbf{j})$ is the $r\times s$-matrix whose (p,q) entry is $A(i_p, j_q)$, $p = 1,\ldots,r$, $q = 1,\ldots,s$. Note the notation $A(i,j)$ for the (i,j)-entry of A used here and below. Also, the `MATLAB` notation $A(:,j)$ for the j-th column of A will be handy.

With $\mathbf{k} := (1,\ldots,k)$, consider the matrix B with entries

$$B(i,j) := \det A(\mathbf{k},i|\mathbf{k},j).$$

On expanding $\det A(\mathbf{k},i|\mathbf{k},j)$ by entries of the last row,

$$B(i,j) = A(i,j)\det A(\mathbf{k}|\mathbf{k}) - \sum_{r\le k} A(i,r)(-)^{k-r}\det A(\mathbf{k}|(\mathbf{k}\backslash r),j).$$

This shows that

$$B(:,j) \in A(:,j)\det A(\mathbf{k}|\mathbf{k}) + \operatorname{span} A(:|\mathbf{k}),$$

while, directly, $B(i,j) = 0$ for $i\in\mathbf{k}$ since then $\det A(\mathbf{k},i|\mathbf{k},j)$ has two rows the same.

In the same way,

$$B(i,:) \in A(i,:)\det A(\mathbf{k}|\mathbf{k}) + \operatorname{span} A(\mathbf{k}|:),$$

while, directly, $B(i,j) = 0$ for $j\in\mathbf{k}$. Thus, if $\det A(\mathbf{k}|\mathbf{k}) \ne 0$, then, for $i > k$,

$$B(i,:)/\det A(\mathbf{k}|\mathbf{k})$$

provides the ith row of the matrix obtained from A after k steps of Gauss elimination (without pivoting). In other words, the matrix $S := B/\det A(\mathbf{k}|\mathbf{k})$ provides the **Schur complement** $S(k+1,\ldots,n|k+1,\ldots,n)$ in A of the **pivot block** $A(\mathbf{k}|\mathbf{k})$.

Since such row elimination is done by elementary matrices with determinant equal to 1, it follows that

$$\det A = \det A(\mathbf{k}|\mathbf{k})\det S(k+1,\ldots,n|k+1,\ldots,n).$$

Since, for any $\#\mathbf{i} = \#\mathbf{j}$, $B(\mathbf{i},\mathbf{j})$ depends only on the square matrix $A(\mathbf{k},\mathbf{i}|\mathbf{k},\mathbf{j})$, this implies

Sylvester's determinant identity. *If*

$$S(i,j) := \det A(\mathbf{k},i|\mathbf{k},j)/\det A(\mathbf{k}|\mathbf{k}), \quad \forall i,j,$$

then

$$\det S(\mathbf{i}|\mathbf{j}) = \det A(\mathbf{k},\mathbf{i}|\mathbf{k},\mathbf{j})/\det A(\mathbf{k}|\mathbf{k}).$$

Cauchy-Binet formula. $\det(AB(\mathbf{i}|\mathbf{j})) = \sum_{\#\mathbf{h}=\#\mathbf{i}} \det(A(\mathbf{i}|\mathbf{h}))\det(B(\mathbf{h}|\mathbf{j}))$.

Carl de Boor, 1989 (latest update: 21oct99)