Determinants

To compute Jordan forms of a given map T, we need to know which scalars c have the property that T-c has a non trivial kernel. It is useful to have a formula in terms of entries of a matrix to determine if the matrix is invertible. One approach is to compute the oriented "volume" of the block spanned by the columns. I.e. this n - dimensional volume would be zero if and only if the columns are dependent. E.g. in a 2 by 2 matrix, the columns are dependent if and only if the parallelogram they span lies in a line, hence has zero area.

This suggests some properties such a formula should have. Since the volume of a block scales by c when we multiply one of the vectors spanning it by c, our volume function ∂(x1,...,xn) should satisfy: ∂(x1, ...,cxi,...,xn) = c ∂ (x1, ...,xi,...,xn). And when we stack two blocks on top of one another by adding two vectors in one entry, the volumes should add. So we should have ∂(x1,...,xi+yi,...,xn) = ∂(x1,...,xi,...,xn) + ∂(x1,...,yi,...,xn). Moreover, when two entries are equal, the vectors spanning the block are dependent, and the block lies in a lower dimensional space, so we should get zero: ∂(x1,...,xi,...,xi,...,xn) = 0. Finally the volume of the unit block spanned by the standard unit vectors should be 1.

So we want a function of n vector variables ∂:(k^n)^n --->k, that has these properties:
i) "alternating": ∂ equals zero when two entries are equal,
ii) "multilinear": ∂ is linear in one variable at a time, and
iii) "normalized": on the block spanned by the standard unit vectors, ∂ has value 1.

Definition: An n dimensional "determinant" is a function ∂ on n by n matrices such that ∂(Id) = 1, and ∂ is alternating, and multilinear as a function say of the columns.

We will show determinant functions exist and are unique, and describe their properties.
Lemma: A determinant function is skew symmetric, i.e. ∂ changes sign when any two entries are exchanged: ∂(x1,...,xi,...,xj,...,xn) = - ∂(x1,...,xj,...,xi,...,,xn).
proof: Since ∂ is alternating, ∂(x1,...,xi+xj,...,xi+xj,...,xn) = 0, and since it is multilinear, this equals 0 = ∂(x1,...,xi,...,xi,...,xn) + ∂(x1,...,xj,...,xj,...,xn) + ∂(x1,...,xi,...,xj,...,xn) + ∂(x1,...,xj,...,xi,...,xn). Since the first two entries are zero, we have our result. **QED.**

Remark: Conversely, skew- symmetric functions are almost alternating, since if ∂(x1,...,xi,...,xj,...,xn) = - ∂(x1,...,xj,...,xi,...,,xn), then 2∂(x1,...,xi,...,xj,...,xn) = 0. This implies ∂(x1,...,xi,...,xj,...,xn) = 0, except in characteristic 2. But over the real field e.g., there is no difference in the two properties, at least for multilinear functions.

We will appeal to a few facts about permutations which we recall. In particular, a permutation of the integers {1,...,n} is a bijective function s:{1,...,n}-->{1,...,n}.
Cor: Permuting the entries of ∂ by a permutation multiplies the value of ∂ by the sign of the permutation.
proof: Recall that the sign of a permutation which exchanges two entries is -1, and the sign of a composition of permutations is the product of their signs. One proof of this is to consider the action of the group of permutations on the ring of polynomials in n variables, Z[X1,...,Xn] by permuting the variables. This action takes the polynomial $\prod$i>j (Xi-Xj)

into itself or minus itself, and we define the sign of the permutation accordingly. The permutation that exchanges X1 and X2, changes the sign of this polynomial since only the sign of (X2-X1) is changed. Similarly every exchange of two adjacent variables changes the sign, and it can be deduced that every exchange of any two variables changes the sign. Since every permutation is a composition of such exchanges, the sign of every composition of permutations is the product of their signs. **QED.**

Uniqueness of determinants

Since a determinant ∂ is multilinear, we regard it as a kind of multiplication. Writing xj = a1j e1 +.....+anj en, in terms of the standard basis, and using multilinearity to expand our product ∂ (x1, x2,..., xn), gives n^n terms. I.e. ∂ (x1, x2,...,xn) = sum over all functions f:{1,..,n}-->{1,..,n} of ∂(a(f(1),1) e(f(1)),..., a(f(n),n) e(f(n))) = sum over all f of ∏j: a(f(j),j) ∂(e(f(1)),...,e(f(n))). Since ∂ = 0 on any sequence with repeated entries, this is really a sum over all bijective functions f, i.e. all permutations s:{1,...,n}-->{1,...,n}.

But since interchanging two entries multiplies the value of ∂ by -1, changing the entries by a permutation s, multiplies the value by sgn(s) = sign of the permutation s. So ∂(x1,...,xn) = Sum over s: sgn(s) ∏j a(s(j),j) ∂(e(1),...,e(n)). Thus the value of a multilinear alternating ∂ is completely determined on every sequence (x1,...,xn), by the value of ∂(e(1),...,e(n)). And for the standard determinant we decreed this value is one. Thus there is at most one way to define a determinant satisfying our three properties, namely by the formula: ∂(x1,...,xn) = Sum over all permutations s: sgn(s) ∏j a(s(j),j).

Define: ∂:Mat nxn (k) --> k, by ∂ ([aij]) = Sum over s: sgn(s) ∏j: a(s(j),j), the only possible multilinear alternating function in the columns of each matrix with ∂(Id) = 1.

The next property is fundamental.
Lemma: ∂(AB) = ∂(A)∂(B), for two nxn matrices A,B.
Proof: Since BA is a linear function of the columns of A, and ∂ is alternating and multilinear in those columns, the composite function d(A) = ∂(BA) is also multilinear and alternating in the columns of A, hence by our uniqueness argument, it is a constant multiple of ∂(A), and the multiplier is d(Id) = ∂(B). Hence ∂(BA) = ∂(B)∂(A). **QED.**

Cor: If A is invertible, then det(A) ≠ 0.
proof: If AB = Id, then det(A)det(B) = det(Id) = 1. **QED.**

Existence of the determinant function.
We must show the function defined by the formula above does have our three properties. If [aij] = Id, all but one term in ∂[Id] is 0, and that one term = 1, so normalization is satisfied. It remains to show this definition of ∂ is indeed multilinear and alternating.

Scaling is the easiest.
Lemma: ∂(x1,...,cxi,...,xn) = c ∂(x1,...,xi,...,xn), for all c in k, all x1,...,xn in k^n.
proof: The definition of ∂ is an alternating sum of terms ∏j: a(s(j),j), and in each of these terms there is one factor a(s(i),i). In the definition of ∂(x1,...,cxi,...,xn), this one factor a(s(i),i) is replaced by c(a(s(i),i)). Thus the whole sum is multiplied by c. **QED.**

Additivity is easy as well.

Lemma: $\partial(x1,\ldots,xi+yi,\ldots,xn) = \partial(x1,\ldots,xi,\ldots,xn) + \partial(x1,\ldots,yi,\ldots,xn)$.

proof: . For simplicity we assume $i = 1$. Again each term of the sum defining ∂, is a product $(a(s(1),1)+b(s(1),1))\prod_{j>1}: a(s(j),j)$, where $y1 =$ the column $[b11\ b21\ldots\ldots bn1]^t$. Since multiplication is distributive, this product is additive in each factor, i.e. $(a(s(1),1)+b(s(1),1))\prod_{j>1}: a(s(j),j) = a(s(1),1)\prod_{j>1}: a(s(j),j) + b(s(1),1))\prod_{j>1}: a(s(j),j)$. Since a sum of additive functions is also additive, we are done. **QED.**

The alternating property follows as well, by grouping terms of the sum.

Lemma: If A has two equal columns, then $\det(A) = 0$.

proof: If t is the involution interchanging say columns 1 and 2, then ts acts the same as s on all other columns but interchanges the value of s at columns 1 and 2, so $a(s(j),j) = a(st(j),j)$ for $j > 2$. But every entry of column 1 equals the corresponding entry of column 2, so $a(s(1),1) = a(s(1),2) = a(st(2),2)$, and $a(s(2),2) = a(st(1),2) = a(st(1),1)$. Thus the two products $\prod_j a(s(j),j)$ and $\prod_j a(st(j),j)$ are equal, but they occur in the sum defining $\det(A)$ with opposite signs since $\mathrm{sgn}(st) = \mathrm{sgn}(t)\mathrm{sgn}(s) = -1\mathrm{sgn}(s)$. Thus every term in the sum for detA is canceled by another term which is equal but of opposite sign. $\det A = 0$. **QED**

Hence $\partial(A)$ is indeed alternating and multilinear in the columns of A. Thus there does exist one and only one function satisfying the properties of a determinant.

Lemma: If $A^t = [aji]$ is the transpose of $A = [aij]$, then $\partial(A^t) = \partial(A)$.

proof: $\partial(A^t) =$ the sum over all s, of $\mathrm{sgn}(s)\prod_j a(j,s(j))$, and if we reorder the factors in this product by the second subscript, this equals $\mathrm{sgn}(s)\prod_j a(s^{-1}(j),j)$. Since s and s^{-1} have the same sign, this is also $\mathrm{sgn}(s^{-1})\prod_j a(s^{-1}(j),j)$. Summing over the inverse of all permutations is the same as summing over all permutations, so this equals the sum over all s, of $\mathrm{sgn}(s)\prod_j a(s(j),j) = \partial(A)$. **QED.**

It is often useful to simplify a matrix by row reduction to introduce more zeroes before computing a determinant. We already know how scaling and interchanging rows affect the determinant, so we observe the third row operation does not change the determinant. Since $\det A = \det A^t$, we have all the same properties for rows as for columns.

Lemma: Adding a scalar multiple of a row to another does not change the determinant.

proof: Suppose we add c times the jth row xj to the ith row xi. Then by additivity, $\partial(x1,\ldots,xi+cxj,\ldots,xj,\ldots,xn) = \partial(x1,\ldots,xi,\ldots,xj,\ldots,xn) + \partial(x1,\ldots,cxj,\ldots,xj,\ldots,xn)$. By scaling and alternating, the last term is zero. **QED**

It is also useful to know how determinants behave on block matrices.

Lemma: If a matrix C has block form as below, then $\det C = \det A.\det B$.

proof: Consider the block matrix C below, where both A and B are square matrices:

$$\begin{vmatrix} A & * \\ 0 & B \end{vmatrix} = C. \text{ Then } \det(C) = \det A.\det B.$$

If B is not invertible then row reduction will introduce a row of zeroes at the bottom, so both B and the whole block matrix have det = 0. If B is invertible, then row operations of the third kind just discussed above, will change the matrix | * | to the zero matrix, since

the rows of B are a basis for the space containing the rows of | * |. This does not change the determinant but changes C into a block matrix of the following form:
| A 0 |
| 0 B |, which is a product matrix,

| A 0 | | A 0 | | I 0 |
| 0 B | = | 0 I | | 0 B |. Thus,

| A * | | A 0 | | A 0 | | I 0 |
det| 0 B| = det| 0 B| = det | 0 I | . det | 0 B | = detA.detB, by LaGrange's formula.

Cor: The determinant of a diagonal, or upper (or lower) diagonal matrix, equals the product of the diagonal entries.

Expansion along one column or row

The next result allows us to compute an nxn determinant as an alternating sum of (n-1)x(n-1) determinants. Suppose we want to expand along the first column. detA = det(A) = (Sum over s): sgn(s)∏j: a(s(j),j). Each term, i.e. each product ∏j: a(s(j),j) with fixed s, has one factor from the first column of A, namely a(s(1),1).
Thus these can be grouped into n subsets according to the value of s(1).
I.e. det(A) = (Sum, all s): sgn(s)∏j a(s(j),j) =

(Sum,s(1)=1), sgn(s)∏j: a(s(j),j)+...+ (Sum,s(1)=n), sgn(s)∏j: a(s(j),j).

Now every term in the first sum with s(1)=1, contains the factor a(1,1), which can be factored out, and similarly for the other sums. Thus we have: det(A) =

a(1,1)(Sum,s(1)=1): sgn(s)∏j≠1 a(s(j),j) +...+ a(n,1)(Sum,s(1)=n): sgn(s)∏j≠1 a(s(j),j).

Here the first partial sum, (Sum,s(1)=1): sgn(s)∏j≠1 a(s(j),j) is just the determinant of the (n-1) by (n-1) matrix obtained by omitting the first row and first column from A. But the second one, (Sum,s(1)=2): sgn(s)∏j≠1 a(s(j),j) is off by a sign, because in the matrix obtained by omitting the first column and second row of A, the n-1 columns are numbered 2,3,4,...,n, while the n-1 rows are numbered 1,3,...,n. Thus the permutation s with s(1) = 2, inducing the "identity" of these index sets takes 2to1, then 3to3, 4to4,..., n to n, and of course 1to2. This makes the sgn of this s = -1 instead of 1. So we have introduce a minus sign. In case s(1) = i, then s induces the "identity" map from the columns 2,3,...,n, to the rows 1,2,...,i-1,i+1,...,n if s(2)=1, s(3)=2,...,s(i) = i-1, and then s(i+1) = i+1,...,s(n)=n. So s differs from the actual identity permutation of {1,...,n} by i-1 transpositions. Thus each term (Sum,s(1)=i): sgn(s)∏j≠1 a(s(j),j), is equal to (-1)^(i+1) times the determinant obtained from A by omitting the 1^{st} column and ith row.

This implies the determinant of A is a dot product of the first column of A with the vector of (±) the (n-1)x(n-1) determinants just described. Namely the element a(i,1) of the ith row in the first column is multiplied by (-1)^(i+1) times the (n-1)x(n-1) determinant of the matrix obtained by eliminating the first column of A and the ith row.

If we denote by Aij the n-1 by n-1 matrix obtained from A by deleting the ith row and jth column, then we have $\det A = a(1,1)\det A11 - a(2,1)\det A21 + \ldots\ldots + (-1)^{(n+1)} \det An1$.

In the same way, the determinant can be computed as a dot product with any column of A. If we use the jth column, the permutation with $s(j)=i$ inducing the "identity" map from the columns $(1,2,..,j-1,j+1,...,n)$ to the rows $(1,2...,i-1,i+1,..,n)$ will differ from the identity on $\{1,...,n\}$ by i-j transpositions, so we can multiply by $(-1)^{(i+j)}$.

Thus for any fixed j: **detA =**
$\mathbf{(-1)^{(1+j)}a(1,j)\det A1j + (-1)^{(2+j)} a(2,j)\det A2j + \ldots\ldots + (-1)^{(n+j)} a(n,j)\det Anj}$.
Since A and its transpose have the same determinant, we can compute $\det A = \det A^t$, by expanding along columns of A^t, i.e. along rows of A. Another approach to this formula is to check it yields a normalized alternating multilinear function, hence = detA.

It follows that if we form a matrix adj(A) whose rows are those (n-1)x(n-1) determinants, with appropriate signs, then the diagonal entries of the matrix product adj(A).A, will all equal det(A). The off diagonal entries moreover are the determinants of matrices having two equal columns, hence zero. We summarize this as follows.

Classical adjoint of a matrix (not the one from spectral theory)

Let Aij be the (n-1) by (n-1) matrix obtained by eliminating from A its ith row and jth column. The number $(-1)^{(i+j)}\det(Aij)$ is called the cofactor of the entry aij of A. Denote by adj(A) the <u>transpose</u> of the matrix of cofactors of A, i.e. the (ij) entry of adj(A) is the cofactor $(-1)^{(j+i)}\det(Aji)$ of the element aji of A. Then we have:

Proposition (LaGrange-Cramer's rule): $A.\text{adj}(A) = \text{adj}(A).A = \partial(A)\text{Id}$.
proof: We defined adj(A) so that $\text{adj}(A).A = \partial(A).\text{Id}$. Then $\partial(A).\text{Id} = \partial(A^t).\text{Id} = (\partial(A^t).\text{Id})^t = (\text{adj}(A^t).A^t)^t = A.(\text{adj}(A^t))^t = A.\text{adj}(A)$. **QED**

Cor: If $\partial(A) \neq 0$, then $(1/\partial(A))\text{adj}(A).A = \text{Id}$, in particular A is invertible.

Cor: (Cayley Hamilton) If $ch(t) = \det(t-A)$ is the characteristic polynomial of A, then $ch(A) = [0]$, i.e. A satisfies its own characteristic polynomial.
proof: By Cramer's rule, $ch(t).\text{Id} = \det(t-A).\text{Id} = \text{adj}(t-A).(t-A)$. Thus in the ring of polynomials with matrix coefficients, the polynomial ch(t).Id is divisible by (t-A) from the right. Hence the value of ch(t) when A is substituted for t from the right is zero, by the non commutative factor theorem, with essentially the same proof as the commutative one. Indeed the coefficients of ch(t).Id are scalars from k hence commute with A, so in any sense, we have $ch(A) = [0]$. **QED.**

Remark: As an example of the non commutative remainder/factor theorem, let $f(t) = Ct^n$. Then right evaluation of f at A equals CA^n, hence $f(t) - f_{right}(A) = C(t^n - A^n)$ is right divisible by (t-A), since $(t^n - A^n) = (t^{n-1} + t^{n-2} A + \ldots + tA^{n-2} + A^{n-1})(t-A)$. Hence $f(t) = f_{right}(A) + g(t)(t-A)$, so by uniqueness of divisibility by the monic (t-A), the remainder after right division by (t-A) equals the right value $f_{right}(A)$. Applying this argument to every term of a general f gives the result.