Determinant (Theory)

This document is a more in-depth discussion of determinant from the point of view of multilinear algebra.

Definition. Let V be a vector space. An n-variable function f on V is *multilinear* if it is linear in each variable when the other variables are fixed, i.e. for each i, we have

$$f(\vec{v}_1, \cdots, \vec{v}_{i-1}, \lambda\vec{v}_i + \mu\vec{v}'_i, \vec{v}_{i+1}, \cdots, \vec{v}_n) = \lambda f(\vec{v}_1, \cdots, \vec{v}_i, \cdots, \vec{v}_n) + \mu f(\vec{v}_1, \cdots, \vec{v}'_i, \cdots, \vec{v}_n)$$

If $\vec{e}_1, \cdots, \vec{e}_m$ is a basis for V, then f is determined by the values $f(\vec{e}_{i_1}, \cdots, \vec{e}_{i_n})$, where $1 \leq i_1, \cdots, i_n \leq m$. This gives a total of $(\dim V)^n$ values, which is sometimes called a tensor in applied math settings, where it is treated as a multi-dimensional generalization of matrices.

Example. If $n = 2$ and $V = \mathbb{R}^m$, then we have a bilinear form, which has the general form

$$f(\vec{v}, \vec{w}) = \vec{v}^T A\vec{w}$$

where A is an $m \times m$ matrix such that $A_{ij} = f(\vec{e}_i, \vec{e}_j)$.

Definition. A multilinear form f is *(totally) symmetric* if its value does not change when the inputs are permuted, i.e.

$$f(\vec{v}_1, \cdots, \vec{v}_i, \cdots, \vec{v}_j, \cdots, \vec{v}_n) = f(\vec{v}_1, \cdots, \vec{v}_j, \cdots, \vec{v}_i, \cdots, \vec{v}_n)$$

It is *(totally) anti-symmetric* or *alternating* if

$$f(\vec{v}_1, \cdots, \vec{v}_i, \cdots, \vec{v}_j, \cdots, \vec{v}_n) = -f(\vec{v}_1, \cdots, \vec{v}_j, \cdots, \vec{v}_i, \cdots, \vec{v}_n)$$

This is equivalent to forcing $f(\vec{v}_1, \cdots, \vec{v}_n) = 0$ if two of the inputs are equal.[1]

Definition. Let V be a vector space of dimension n. A *volume form* is an alternating multilinear n-form.

Theorem. All volume forms on V are proportional to each other.

Proof. Let ω be a volume form on V. Let $\vec{v}_1, \cdots, \vec{v}_n$ be a basis of V, then by multilinearity, ω is determined by its values $\omega(\vec{v}_{i_1}, \cdots, \vec{v}_{i_n})$. Since ω is alternating, this is zero if two of the indices agree, so $i_1, \cdots, i_n$ is a permutation of $1, \cdots, n$. But by anti-symmetry, we have

$$\omega(\vec{v}_{i_1}, \cdots, \vec{v}_{i_n}) = (-1)^{\mathrm{sign}(i_1, \cdots, i_n)} \omega(\vec{v}_1, \cdots, \vec{v}_n)$$

It follows that ω is uniquely determined by its value $\omega(\vec{v}_1, \cdots, \vec{v}_n)$. □

Theorem. Let ω be a volume form on V, and let $T : V \to V$ be a linear transformation, then the function $T^*\omega$ defined by

$$T^*\omega(\vec{v}_1, \cdots, \vec{v}_n) = \omega(T(\vec{v}_1), \cdots, T(\vec{v}_n))$$

is also a volume form on V.

Proof. This follows from the axioms of a volume form. □

By the previous two theorems, we can make the following definition.

Definition. The *determinant* of a linear transformation $T : V \to V$ is the number $\det(T)$ such that

$$T^*\omega = (\det T)\omega$$

for any volume form ω on V.

Corollary. If T and S are two linear transformations $V \to V$, then

$$\det(T \circ S) = \det(T)\det(S)$$

[1] We are working with scalars $\mathbb{R}$ or $\mathbb{C}$. It may not be true in general.

Proof. This is immediate from the observation that $(T \circ S)^*\omega = S^*T^*\omega$. $\square$

Suppose now $V = \mathbb{R}^n$, so it has a standard basis $\vec{e}_1, \cdots, \vec{e}_n$. There are now two ways to view the determinant of an $n \times n$ matrix:

1. An $n \times n$ matrix is a linear transformation from $\mathbb{R}^n$ to itself, so we can use the above definition of the determinant of a linear transformation.

2. A sequence of n vectors form an $n \times n$ matrix, so the determinant can be defined as the unique volume form such that $(\vec{e}_1, \cdots, \vec{e}_n)$ is sent to 1.

The usual properties of determinants, especially $\det(AB) = \det(A)\det(B)$ follow immediately from this definition. The column expansion formula is also easy to show. The difficulty property to prove from this point of view is

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

This is difficult because we haven't introduced a coordinate-free way of looking at the transpose, which comes from the dual space. Instead, we will use the formula for determinant in terms of permutations.

We now present a generalization of $\det(AB) = \det(A)\det(B)$.

Theorem (Binet–Cauchy formula)**.** Let A be an $n \times p$ matrix and B be a $p \times n$ matrix, then

$$\det(AB) = \sum_{1 \le i_1 < \cdots < i_n \le p} \det(A[1, \cdots, n; i_1, \cdots, i_n]) \det(B[i_1, \cdots, i_n; 1 \cdots n])$$

where $A[1, \cdots, n; i_1, \cdots, i_n]$ denote the $n \times n$ minor of A defined by taking columns $i_1, \cdots, i_n$, and similarly for $B[i_1, \cdots, i_n; i, \cdots, n]$.

Proof. We view A and B as linear transformations. Let ω be a volume form on $\mathbb{R}^n$, then $A^*\omega$ is an alternating n-form on $\mathbb{R}^p$, which is determined by its values

$$A^*\omega(\vec{e}_{i_1}, \cdots, \vec{e}_{i_n}) = \det(A[1, \cdots, n; i_1, \cdots, i_n])$$

over all $1 \le i_1 < \cdots < i_n \le p$. Using this formula to compute $B^*A^*\omega$ gives the right hand side of the theorem times ω. But this is $(AB)^*\omega$, which is by definition $\det(AB)\omega$. $\square$

Example. Let $A = \left(\begin{smallmatrix} a_1 & \cdots & a_n \\ b_1 & \cdots & b_n \end{smallmatrix}\right)$, then

$$AA^T = \begin{pmatrix} \sum a_i^2 & \sum a_i b_i \\ \sum a_i b_i & \sum b_i^2 \end{pmatrix}$$

so $\det(AA^T) = (\sum a_i^2)(\sum b_i^2) - (\sum a_i b_i)^2$. On the other hand, by the Binet–Cauchy formula, this is

$$\det(AA^T) = \sum_{1 \le i < j \le n} \det(A[1, 2; i, j]) \det(A[1, 2; i, j]^T) = \sum_{1 \le i < j \le n} (a_i b_j - a_j b_i)^2$$

This gives the identity used in our proof of the Cauchy–Schwarz inequality.

Volume

This section tries to justify calling the determinant a volume form. The volume should be a function $\text{vol}(\cdot)$ whose input is a subset of $\mathbb{R}^n$ and outputs a non-negative real number. It should also satisfy

1. Additivity: If $A \cap B = \emptyset$, then $\text{vol}(A \cup B) = \text{vol}(A) + \text{vol}(B)$.

2. Translation-invariance: If $\vec{v} \in \mathbb{R}^n$, then $\text{vol}(\vec{v} + A) = \text{vol}(A)$.

3. Normalization: The unit cube $[0, 1]^n$ has volume 1.

It is true that *something like* the first two properties characterizes volume up to a constant multiple, and it is completely determined when the third property is added. We will pretend that this is true, so we do not have to introduce measure theory.

Now let T be a linear transformation of $\mathbb{R}^n$. The function $S \mapsto \text{vol}(T^{-1}(S))$ also satisfies additivity and translation-invariance, so it is a multiple of the volume, i.e. there exists a constant c_T such that $\text{vol}(T(S)) = c_T \text{vol}(S)$ for all $S \subseteq \mathbb{R}^n$. We want to show that $c_T = |\det T|$. The sign of the determinant keeps track of the orientation, which is a whole other story.

Both sides of the equation are multiplicative, so by the singular value decomposition, we just need to prove it for orthogonal transformations and scaling in one coordinate direction.

- Orthogonal case: by definition, T preserves the unit ball $B = \{\vec{x} \in \mathbb{R}^n | \|\vec{x}\| \leq 1\}$. Setting $S = B$ in the above formula gives $\text{vol}(B) = c_T \text{vol}(B)$, so $c_T = 1$ (we also need $\text{vol}(B) \neq 0$, which is not hard to show from the axioms).

- Scaling: let T be the transformation which scales the first entry by $\alpha > 0$. Let C be the unit cube, then $T(C)$ is the cube $C_\alpha = [0, \alpha] \times [0, 1]^{n-1}$. By additivity, and translation invariance, we have

  $$\text{vol}(C_{\alpha+\beta}) = \text{vol}(C_\alpha) + \text{vol}(C_\beta)$$

  Therefore, $\text{vol}(C_\alpha)$ is an increasing, additive function in α. All such functions have the form $c\alpha$ for some constant c, which we know must be 1 since $\text{vol}(C_1) = 1$ by normalization. This shows that $c_T = \text{vol}(C_\alpha) = \alpha = |\det T|$.

Cayley–Hamilton theorem

Theorem. Let A be an $n \times n$ matrix. Let $p(\lambda)$ be its characteristic polynomial, so

$$p(\lambda) = \det(\lambda I_n - A)$$

then $p(A) = 0$.

Proof. Let $B = \text{adj}(\lambda I_n - A)$, then each entry of B is the determinant of an $(n-1) \times (n-1)$ matrix, each entry of which is a polynomial in λ of degree at most 1, so it is a polynomial of degree at most $n-1$, and we can write

$$B = B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1}$$

where $B_0, B_1, \cdots, B_{n-1}$ are $n \times n$ matrices.

Using the property of the adjugate matrix, we have

$$\begin{aligned} \det(\lambda I_n - A)I_n &= (B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1})(\lambda I_n - A) \\ &= -B_0 A + \sum_{i=1}^{n-1}(B_{i-1} - B_i A)\lambda^i + B_{n-1}\lambda^n \end{aligned}$$

Suppose $p(\lambda) = \sum_{i=0}^{n} c_i \lambda^i$, then we have

$$B_{n-1} = c_n I_n,\ -B_0 A = c_0 I_n,\ B_{i-1} - B_i A = c_i I_n \text{ for } i = 1, 2, \cdots, n-1$$

Multiply the equation involving c_i by A^i on the right and sum:

$$\begin{aligned} p(A) &= c_0 I_n + c_1 A + \cdots + c_n A^n \\ &= -B_0 A + (B_0 - B_1 A)A + (B_1 - B_2 A)A^2 + \cdots + (B_{n-2} - B_{n-1}A)A^{n-1} + B_{n-1}A^n \\ &= (-B_0 A + B_0 A) + (-B_1 A^2 + B_1 A^2) + \cdots + (-B_{n-1}A^n + B_{n-1}A^n) \\ &= 0 \end{aligned}$$

This finishes the proof. ∎