5. Determinant and trace.

The concept of the determinant of a linear map is simple - we will give an unofficial geometric definition as motivation, and then formalize it.

Def: (Unofficial). Let \( T: \mathbb{R}^n \rightarrow \mathbb{R}^n \) be a linear map, and \([0,1]^n\) the unit n-cube. The determinant of \( T \) is the "signed volume" of \( T([0,1]^n) \).

Our unofficial definition has some interesting properties:

- \( \det T = 0 \) iff \( T \) is not invertible
- \( \det (TS) = (\det T)(\det S) \)
- If \( T \) and \( S \) differ by swapping two columns, then \( \det T = -\det S \).
- \( \det T \) is "linear" in each column - i.e., if \( T \) is obtained from \( S \) by multiplying a column by \( c \), then \( \det T = c(\det S) \).
**Permutations:**

**Def:** Let $[n] := \{1, \ldots, n\}$. A permutation is a bijection $\pi : [n] \rightarrow [n]$. The set of all permutations of $n$ elements is denoted $S_n$, and is a group.

We can describe a permutation by a table, or more concisely, by cycle notation:

**Example:** $\pi : \{1, 2, 3, 4\} \rightarrow \{1, 2, 3, 4\}$.

**Table notation:**

\[
\begin{array}{c|cccc}
  i & 1 & 2 & 3 & 4 \\
  \pi(i) & 2 & 4 & 1 & 3 \\
\end{array}
\]

**Cycle notation:** $\pi = (1 \ 2 \ 4 \ 3)$ meaning $1 \rightarrow 2 \rightarrow 4 \rightarrow 3$.

$\pi^2 = (1 \ 4)(2 \ 3)$, $\pi^3 = (1 \ 3 \ 4 \ 2)$, $\pi^4 = (1)(2)(3)(4)$.

By convention, we usually omit length-1 cycles, e.g., $(1)(2)(3)(4)$.

**Def:** Let $x_1, \ldots, x_n$ be $n$ variables. Their discriminant is defined as $P(x_1, \ldots, x_n) = \prod_{i < j} (x_i - x_j)$.

For a permutation $\pi \in S_n$,

\[
P(\pi(x_1, \ldots, x_n)) = \prod_{i < j} (x_{\pi(i)} - x_{\pi(j)}) = \pm P(x_1, \ldots, x_n).
\]
Def: The **signature** \( \text{sgn}(\pi) \) of a permutation \( \pi \in S_n \) is defined as \( \text{P}(\pi(x_1, \ldots, x_n)) = \text{sgn}(\pi) \text{P}(x_1, \ldots, x_n) \).

Clearly, \( \text{sgn}(\pi) = \pm 1 \).

Def: A **transposition** is a permutation \( \tau \in S_n \) such that

For some \( j \neq k \in [n] \),
\[
\begin{align*}
\tau(i) &= i \quad i \neq j, k \\
\tau(j) &= k \\
\tau(k) &= j.
\end{align*}
\]

**Proof:**

(i) \( \text{sgn}(\pi_1 \circ \pi_2) = \text{sgn}(\pi_1) \text{sgn}(\pi_2) \)

(ii) \( \text{sgn}(\tau) = -1 \) for any transposition

(iii) Every permutation \( \pi \in S_n \) can be written as a composition of transpositions: \( \pi = \tau_k \circ \cdots \circ \tau_1 \).

(iv) This decomposition is not unique, but the parity of \( k \), the number of transpositions is.

(v) If \( \pi = \tau_{k_1} \circ \cdots \circ \tau_1 \), then \( \text{sgn}(\pi) = (-1)^k \).

**Proof:** Exercise.

**Multilinear forms**

**Def:** A **k-linear form** is a function \( f: X_1 \otimes \cdots \otimes X_k \to K \)

(we'll assume \( X_1 = \cdots = X_k = X \)) that is linear in each coordinate, i.e., upon fixing any \( k-1 \) arguments, it remains linear in the remaining argument.
Examples:

1. 1-linear forms are just functions.
2. 2-linear forms are bilinear forms.
3. A 3-linear form \( f : X \otimes X \otimes X \to K \) has identities such as
   \[ f(a_1x_1 + a_2x_2, y, z) = a_1f(x_1, y, z) + a_2f(x_2, y, z), \]
   and similarly for \( f(x, a_1y_1 + a_2y_2, z), f(x, y, a_1z_1 + a_2z_2), \) etc.

Theorem 5.1: The set of \( k \)-linear forms is a vector space of dimension \( n^k \).

Proof: (sketch). Verify that a basis consists of functions

\[ \{ F_{j_1 \ldots j_k}(x_{i_1}, \ldots, x_{i_k}) = \delta_{j_1 i_1} \ldots \delta_{j_k i_k} : 1 \leq j_k \leq n \} \]

For any permutation \( \pi \in S_k \), write

\[ (\pi f)(x_{i_1}, \ldots, x_{i_k}) = f(x_{\pi(i_1)}, \ldots, x_{\pi(i_k)}). \]

Note that for any \( k \)-linear form \( f \), \( \pi f \) is also \( k \)-linear.

DEF: A \( k \)-linear form is symmetric if \( \pi f = f \) for all \( \pi \in S_k \).

Examples:

1. \( f(x_1, x_2) = l_1(x_1)l_2(x_2) + l_1(x_2)l_2(x_1) \) for fixed \( l_1, l_2 \in \mathcal{X} \).
2. \( f(x_1, \ldots, x_k) = \sum_{\pi \in S_k} \pi f(x_{\pi(1)}, \ldots, x_{\pi(k)}) \)
Def: A $k$-linear form is skew-symmetric if $\tau f = -f$ for every transposition $\tau \in S_k$.

Example: $f(x_1, x_2) = l_1(x_1) l_2(x_2) - l_1(x_2) l_2(x_1)$.

Def: A $k$-linear form is alternating if $f(x_1, \ldots, x_k) = 0$ if $x_i = x_j$ for some $i \neq j$.

Prop: The set of alternating (resp. symmetric, or skew-symmetric) $k$-linear forms is a subspace.

Proof: Exercise.

Theorem 5.2: Every alternating multilinear form is skew-symmetric.

Proof: Pick $i \neq j$, and define $g(x_i, x_j) = f(x_1, \ldots, x_k)$ (i.e., the other entries are fixed). Note that $g$ is bilinear, alternating.

Thus, $0 = g(x_i + x_j, x_i + x_j) = g(x_i, x_i) + g(x_j, x_j) + g(x_i, x_j) + g(x_j, x_i)$

$= g(x_i, x_j) + g(x_j, x_i)$.

$\Rightarrow g(x_i, x_j) = -g(x_j, x_i)$

$\Rightarrow \tau f = -f$ for $\tau = (i \ j)$. $\square$

Remark: The converse "almost" holds.

Suppose $f = -f$. Then $(1+1)f = 0 \Rightarrow f = 0$ or $1+1 = 0$ (e.g., if $K = \mathbb{Z}_2$). Thus, the converse holds for $K \neq \mathbb{Z}_2$.
Theorem 5.3: If \( x_1, \ldots, x_k \) are linearly dependent and \( f \) is an alternating \( k \)-linear form, then \( f(x_1, \ldots, x_n) = 0 \).

Proof: If \( x_i = 0 \), the result is trivial. Otherwise, write \( x_h = \sum_{i=1}^{k-1} a_i x_i \) for some \( x_h \), and replace \( x_h \) in \( f(x_1, \ldots, x_k) \) by this sum.

Use linearity to break apart \( f(x_h, \ldots, x_k) \) into a sum, each one having some \( x_j \) in the \( h \)-position, \( j \neq h \). All of these terms are zero, since \( f \) is alternating. \( \square \)

The converse holds in one important special case: \( k=n \).

Theorem 5.4: If \( f \) is a non-zero alternating \( n \)-linear form, and \( x_1, \ldots, x_n \) are linearly independent, then \( f(x_1, \ldots, x_n) \neq 0 \).

Proof: By assumption, \( x_1, \ldots, x_n \) is a basis for \( X \).

Pick \( y_1, \ldots, y_n \in X \) and write \( y_i = a_{i1} x_1 + \cdots + a_{in} x_n \).

Now, \( f(y_1, \ldots, y_n) = f \left( \sum a_{i1} x_1, \ldots, \sum a_{in} x_n \right) \)

\[
= \sum f(z_1, \ldots, z_n) \quad \text{where } z_j \text{ distinct } x_i's.
\]

\[
= \sum f(x_{\pi(1)}, \ldots, x_{\pi(n)})
\]

\[
= \sum \text{sgn} (\pi) f(x_1, \ldots, x_n).
\]

If \( f(x_1, \ldots, x_n) = 0 \) then \( f(y_1, \ldots, y_n) = 0 \) \( \Rightarrow \) \( f = 0 \). \( \square \)
Remark: The proof of Theorem 5.4 yields an important corollary:

Theorem 5.5: Any two alternating \(n\)-linear forms are linearly dependent.

Proof: Let \(f, g\) be alternating \(n\)-linear forms, and let \(x_1, \ldots, x_n\) be a basis of \(X\).

Pick \(y_1, \ldots, y_n \in X\).

As before, \(f(y_1, \ldots, y_n) = \sum_{\text{lots}} \text{sgn}(\pi) f(x_1, \ldots, x_n) = a \in K\).

\(g(y_1, \ldots, y_n) = \sum_{\text{lots}} \text{sgn}(\pi) g(x_1, \ldots, x_n) = b = \lambda a \in K\).

Thus, \(g(x_1, \ldots, x_n) = \lambda f(x_1, \ldots, x_n)\). \(\square\)

Thus, the space of alternating \(n\)-linear forms is at most one-dimensional.

Theorem 5.6: The vector space of alternating \(n\)-linear forms is one-dimensional.

Proof: It suffices to show that there exists one such form. We'll use induction on \(k \leq n\).

Base case \((k=1)\): Any \(l \neq 0\) in \(X'\) is a non-trivial \(1\)-linear form.

Now, suppose \(F\) is a non-zero alternating \(k\)-linear form, for \(k < n\). We will construct a non-zero alternating \((k+1)\)-linear form \(g\).
Pick $y_1, \ldots, y_k$ such that $f(y_1, \ldots, y_k) \neq 0$, and find $y_{k+1} \in X$ not in the subspace $Y := \text{span}\{y_1, \ldots, y_k\}$.

Pick $l \in Y^*$ such that $l(y_{k+1}) \neq 0$, and $l(y_1) = \cdots = l(y_k) = 0$.

Define: $g(x_1, \ldots, x_k, x_{k+1}) = -f(x_1, \ldots, x_k)l(x_{k+1}) + \sum_{i=1}^{k} (i \cdot l(x_i)) f(x_1, \ldots, x_k)l(x_{k+1})$

[For example, if $k = 3$, then $g(y_1, x_1, x_2, x_3) = f(x_1, y_3, x_3)l(x_3) + f(x_1, x_2, x_3)l(x_2)$ $+ f(x_1, x_2, x_3)l(x_3) - f(x_1, x_2, x_3)l(x_4)$.

It is elementary to prove that this is $k$-linear. We must show that it is non-zero and alternating.

**Non-zero:** $g(y_1, \ldots, y_k, y_{k+1}) = -f(y_1, \ldots, y_k)l(y_{k+1}) \neq 0$.

**Alternating:** Consider vectors $x_1, \ldots, x_k, x_{k+1}$ such that $i < j$ but $x_i = x_j$. It suffices to prove $g(x_1, \ldots, x_k, x_{k+1}) = 0$.

Note: $x_i$ and $x_j$ occur in all but two "terms" of $g$. All other (besides these two) terms vanish.

**Case 1:** $(j = k+1)$:

$g(x_1, \ldots, x_{k+1}) = (i \cdot k+1) f(x_1, \ldots, x_k)l(x_{k+1}) - f(x_1, \ldots, x_k)l(x_{k+1})$

$= 0$ (because $x_i = x_{k+1}$).

**Case 2:** $(j \leq k)$:

$g(x_1, \ldots, x_{k+1}) = \left[ (i \cdot j) f(x_1, \ldots, x_k) - f(x_1, \ldots, x_k) \right] l(x_{k+1})$

$= 0$ (because $f$ is alternating).
let $f$ be an alternating $n$-linear form on $X$, and $T : X \to X$ a linear map.

Note that $(\bar{T}f)(x_1, \ldots, x_n) := f(Tx_1, \ldots, Tx_n)$ is alternating and $n$-linear as well.

Moreover, $\bar{T}$ is a linear map on the (one-dimensional) space of all such forms, thus, $\bar{T} : f \mapsto \lambda f$ for some $\lambda \in K$.

Def: This scalar $\lambda$ is the determinant of $T$, denoted $\det T$.

Thus, the determinant satisfies the following:

Universal property: Given a linear map $T : X \to X$, there exists a unique scalar $\lambda \in K$ such that for every alternating $n$-linear form $f$,

$$ f(Tx_1, \ldots, Tx_n) = \lambda f(x_1, \ldots, x_n). $$

\[ \begin{array}{c}
\{V^n\} & \xrightarrow{T_0 \ldots \otimes T} & \{V^n\} \\
\downarrow f & & \downarrow f \\
\{K\} & \xrightarrow{\lambda} & \{K\}
\end{array} \]

Remark: This definition is not only independent of matrices, but also independent of choice of basis!

Example: If $Tx = cx$, then

$$(\bar{T}f)(x_1, \ldots, x_n) = f(cx_1, \ldots, cx_n) = c^n f(x_1, \ldots, x_n).$$

Thus, $\det T = c^n$. It follows that $\det 0 = 0$, $\det I = 1$. 
**Theorem 5.7:** For any two linear maps $A, B : X \to X$,

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

**Proof:** Write $C = AB$. For an alternating $n$-linear form $f$,

$$(Cf)(x_1, \ldots, x_n) = f(ABx_1, \ldots, ABx_n) = (Af)(Bx_1, \ldots, Bx_n) = (ABf)(x_1, \ldots, x_n).$$

Thus, $C = BA$. We have $Cf = (\det C)f$ and $Af = (\det A)f$.

$$= \det(AB) = (\det A)(\det B). \quad \square$$

**Corollary:** $A : X \to X$ is invertible iff $\det A \neq 0$.

**Proof:** If $A$ is invertible, then $AA^{-1} = I$.

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

Conversely, if $\det A \neq 0$, and $x_1, \ldots, x_n$ is a basis of $X$, and $f$ a non-zero alternating $n$-linear form on $X$, then

$$(\det A)f(x_1, \ldots, x_n) \neq 0 \quad (by \, Theorem \, 5.4)$$

$$\Rightarrow \{Ax_1, \ldots, Ax_n\} \text{ is linearly independent} \quad (by \, Theorem \, 5.3)$$

$$\Rightarrow A \text{ is invertible}. \quad \square$$
Determinants and matrices

Let $f$ be a non-zero alternating $n$-linear form.

Let $\{x_1, \ldots, x_n\}$ be a basis of $X$, and let $A = (a_{ij})$ be the matrix of a linear map $X \to X$ with this basis.

Recall that $A x_j = \sum_{i=1}^n a_{ij} x_i$.

Thus, $(\det A) f(x_1, \ldots, x_n) = f(A x_1, \ldots, A x_n)$

$$= f\left(\sum_{i=1}^n a_{i1} x_1, \ldots, \sum_{i=1}^n a_{in} x_i\right)$$

Now, expand this by multilinearity, to get a large sum of terms such as $f(z_1, \ldots, z_n)$ where each $z_i$ is a distinct $x_i$ (if $z_i = z_j$ then $f(z_1, \ldots, z_n) = 0$).

Moreover, every permutation $f(x_{\pi(1)}, \ldots, x_{\pi(n)}) = \prod f(x_i, \ldots, x_n)$ appears in this sum.

**Remark.** The coefficient of $\prod f(x_i, \ldots, x_n)$ is $a_{\pi(1),1} \cdots a_{\pi(n),n}$.

Since $f$ is skew-symmetric, we have

$$\det A = \sum_{\pi \in S_n} (\text{sgn } \pi) a_{\pi(1),1} \cdots a_{\pi(n),n} \quad (\ast)$$

Also note that if $\pi, \sigma \in S_n$, then $a_{\pi(1),1} \cdots a_{\pi(n),n}$ and $a_{\sigma(1),1} \cdots a_{\sigma(n),n}$ differ in the order of the factors only.
Let's apply this with $\sigma = \pi^{-1}$. (Note: $\text{sgn}(\pi) = \text{sgn}(\pi^{-1})$)

$$\det A = \sum_{\pi \in S_n} (\text{sgn} \pi) a_{\pi(1)} \cdots a_{\pi(n)} = \sum_{\pi \in S_n} (\text{sgn} \pi) a_{1, \pi(1)} \cdots a_{n, \pi(n)} = \det A^T.$$ 

In summary, the determinant of $A$ can be thought of as an $n$-linear function of its column vectors, or of its row vectors. In fact it is the unique "normalized" alternating $n$-linear form, in that $f(e_1, \ldots, e_n) = 1$, ($\det I = 1$).

**Lemma 5.8**: Let $A = (c_1, \ldots, c_n)$ be a matrix ($c_i$ is a column vector) and let $B$ be the matrix obtained from $A$ by adding $k$ times the $i^{th}$ column of $A$ to the $j^{th}$ column, $i \neq j$.

Then $\det A = \det B$.

**Proof**: Exercise.

**Lemma 5.9**: Let $A$ be an $n \times n$ matrix whose first column is $e_1$.

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

Then $\det A = \det A_{11}$.

**Proof**: By lemma 5.8, $\det A = \det \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}$.

Define $f(A_{11}) = \det \begin{pmatrix} 1 & 0 \\ 0 & A_{11} \end{pmatrix}$, a function of an $(n-1) \times (n-1)$ matrix $A_{11}$.
Clearly, $f$ is an alternating $n$-linear form with $f(I) = 1$.
so it must be the determinant function.

**Corollary 5.10**: Let $A$ be a matrix whose $j^{th}$ column is $e_j$.
Then $\det A = (-1)^{i+j} \det A_{ij}$, where $A_{ij}$ is the $(n-1) \times (n-1)$
matrix obtained by striking out the $i^{th}$ row and $j^{th}$ column
of $A$ (called the $(ij)^{th}$ minor of $A$).

**Proof**: Exercise.

**Theorem 5.11**: Let $A$ be an $n \times n$ matrix, and $1 \leq j \leq n$.

Then $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} (\det A_{ij})$. (The "Laplace expansion").

**Proof**: To simplify notation, take $j = 1$.

Write $a_i = a_i e_i + \ldots + a_n e_n$ where $A = (a_1, \ldots, a_n)$.

By multilinearity, $\det A = f(a_1, \ldots, a_n)$

$= f(a_1 e_1 + \ldots + a_n e_n, a_2, \ldots, a_n)$

$= a_1 f(e_1, a_2, \ldots, a_n) + \ldots + a_n f(e_n, a_2, \ldots, a_n)$

Now apply Corollary 5.10. □

**Application**: Systems of equations.

Consider a system $A x = u$, $A$ is invertible and $x = \sum_{j=1}^{n} x_j e_j$,

$A = (a_1, \ldots, a_n) = (A e_1, \ldots, A e_n)$. 
Thus, our system can be written as \( \sum_{j=1}^{n} x_j a_j = u \).

Let \( A_k = (a_1, \ldots, a_{k-1}, u, a_{k+1}, \ldots, a_n) \) for each \( k \).

\[
A_k = (a_1, \ldots, a_{k-1}, \sum_{j=1}^{n} x_j a_j, a_{k+1}, \ldots, a_n).
\]

By multilinearity, \( \det A_k = \sum_{j=1}^{n} x_j \det (a_1, \ldots, a_{k-1}, a_j, a_{k+1}, \ldots, a_n) \)

\[
= x_k \det (a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_n)
\]

\[
= x_k \det A.
\]

Since \( A \) is invertible, \( \det A \neq 0 \) \( \Rightarrow \) \( x_k = \frac{\det A_k}{\det A} \).

Now apply the Laplace Expansion (Thm 5.11) to \( A_k \):

\[
\det A_k = \sum_{i=1}^{n} (-1)^{i+k} \det A_{ik} u_i.
\]

\[
\Rightarrow x_k = \sum_{i=1}^{n} (-1)^{i+k} \frac{\det A_{ik}}{\det A} u_i. \quad \text{This is "Cramer's rule."}
\]

**Theorem 5.12**: If \( A \) is invertible, then its inverse \( A^{-1} \)

has the form \( (A^{-1})_{ki} = (-1)^{i+k} \frac{\det A_{ik}}{\det A} \).

**Proof**: Consider a system \( AX = u \) \( \Rightarrow \) \( X = A^{-1}u \). By Cramer's rule:

\[
\sum_{i=1}^{n} (-1)^{i+k} \frac{\det A_{ik}}{\det A} u_i = x_k = (A^{-1}u)_k = \sum_{i=1}^{n} (A^{-1})_{ki} u_i.
\]

**Note**: This formula isn't "practical" for computing inverses.
Def: The trace of an $n \times n$ matrix is $\text{tr } A = \sum_{i=1}^{n} a_{ii}$.

Theorem 5.13:

(a) Trace is linear: $\text{tr}(k A) = k \text{tr } A$,

$$\text{tr}(A+B) = \text{tr } A + \text{tr } B.$$ 

(b) Trace is "commutative": $\text{tr}(AB) = \text{tr}(BA)$

Proof: (a) is obvious.

For (b): $(AB)_{ii} = \sum_{k=1}^{n} a_{ik} b_{ki}$ and $(BA)_{ii} = \sum_{k=1}^{n} b_{ik} a_{ki}$

Thus, $\text{tr } (AB) = \sum_{i,k} a_{ik} b_{ki} = \sum_{i,k} b_{ik} a_{ki} = \text{tr } (BA)$. \[ \square \]

Theorem 5.14: Similar matrices have the same determinant and trace.

Proof: Suppose that $A = S B S^{-1}$.

$$\det A = \det(SBS^{-1}) = (\det S)(\det B)(\det S^{-1}) = (\det B)(\det S)(\det S^{-1})$$

$$= (\det B) \det (SS^{-1}) = \det B.$$ \[ \checkmark \]

$$\text{tr } A = \text{tr } (SBS^{-1}) = \text{tr } (S^{-1}BS) = \text{tr } B.$$ \[ \checkmark \]

Since similar matrices represent the same linear map but with a different choice of basis, it is well-founded to speak of the determinant and trace as functions of linear maps, not just matrices.