Linear Algebra Notes

Lecture Notes, University of Toronto, Fall 2016

1. Determinants

1.1. **The inverse of a 2×2-matrix.** For a 2×2-matrix $A \in M_{2\times 2}(F)$, given as

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we define its *determinant* by the formula

$$\det(A) = ad - bc.$$

Its importance can be seen from the following

Lemma 1.1. *The 2×2-matrix A is invertible if and only if $\det(A) \neq 0$. In this case, the inverse is given by*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof. Let

$$B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

By carrying out the matrix multiplication, we see that

$$AB = \det(A)\, I$$

where I is the identity matrix. If $\det(A) \neq 0$, this verifies that $\det(A)^{-1}B$ is a matrix inverse of A. If $\det(A) = 0$, the identity becomes $AB = 0$. If A were invertible, then this would give $B = A^{-1}(AB) = A0 = 0$. Hence, all matrix entries $d, -b, -c, a$ of B are zero, which means that $A = 0$, a contradiction. So, A cannot be invertible. □

Note: This is a formula that you should (and I'm sorry to say this) **memorize!!!** Namely: $A^{-1} = \det(A)^{-1}B$; *to get B from A, switch the diagonal entries and put minus signs for the off-diagonal ones.*

Example 1.2. **Problem**: Solve the system of equations

$$2x_1 + 3x_2 = 4$$
$$2x_1 + x_2 = 3$$

Solution: Invert the coefficient matrix, and apply to the column vector on the right side:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \frac{1}{-4} \begin{pmatrix} 1 & -3 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = -\frac{1}{4} \begin{pmatrix} -5 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{5}{4} \\ \frac{1}{2} \end{pmatrix}$$

so $x_1 = \frac{5}{4},\ x_2 = \frac{1}{2}$.

1.2. **Interpretion of the determinant.** What's the meaning of the mysterious expression $\det(A) = ad - bc$? Consider temporarily the case $F = \mathbb{R}$. Let $v_1, v_2 \in \mathbb{R}^2$ be vectors v_1, v_2, and

$$\mathrm{vol}(v_1, v_2) \in \mathbb{R}$$

the *signed area* of the parallelogram spanned by the two vectors. (We write vol, since we will soon generalize to higher dimensions, where one speaks of 'volume') Here the sign is taken to be positive if the positively oriented angle from v_1 to v_2 is between 0 and π, and negative if it is between π and 2π. The following facts are known (mostly from high school geometry).

P1. $\mathrm{vol}(av_1, v_2) = a\,\mathrm{vol}(v_1, v_2) = \mathrm{vol}(v_1, av_2)$,
P2. $\mathrm{vol}(v_1 + av_2, v_2) = \mathrm{vol}(v_1, v_2) = \mathrm{vol}(v_1, v_2 + a\,\mathrm{vol}\,v_2)$,

for all vectors v_1, v_2 and scalars a. Note that this implies $\mathrm{vol}(v_1, v_2) = 0$ if one of v_1, v_2 is zero, and also

$$\mathrm{vol}(v, v) = 0, \quad v \in V$$

by taking $v_1 = 0, v_2 = v, a = 1$ in the second property. Furthermore, we can derive:

Lemma 1.3. *The map* $\mathrm{vol}\colon \mathbb{R}^2 \times \mathbb{R}^2 \to R$ *is bi-linear (i.e., linear in each arguments separately)*

Proof. We have to show that

$$\mathrm{vol}(v_1 + v_1', v_2) = \mathrm{vol}(v_1, v_2) + \mathrm{vol}(v_1', v_2)$$

for all vectors v_1, v_1', v_2. If $v_2 = 0$ this is clear, and if v_1 or v_1' is a multiple $a\,v_2$ it follows from P2. Thus, we may assume that v_1, v_2 are a basis. Write $v_1' = \lambda v_1 + \mu v_2$, and simplify

$$\begin{aligned}\mathrm{vol}(v_1 + v_1', v_2) &= \mathrm{vol}((1+\lambda)v_1 + \mu v_2, v_2)\\ &= \mathrm{vol}((1+\lambda)v_1, v_2)\\ &= (1+\lambda)\,\mathrm{vol}(v_1, v_2)\\ &= \mathrm{vol}(v_1, v_2) + \mathrm{vol}(v_1', v_2).\end{aligned}$$

Thus vol is linear in the first argument , similarly it's also linear in the second argument. □

Remark 1.4. Using the bi-linearity, together with P1 we also see now that

$$0 = \mathrm{vol}(v_1+v_2, v_1+v_2) = \mathrm{vol}(v_1, v_1)+\mathrm{vol}(v_2, v_2)+\mathrm{vol}(v_1, v_2)+\mathrm{vol}(v_2, v_1) = \mathrm{vol}(v_1, v_2)+\mathrm{vol}(v_2, v_1)$$

thus

$$\mathrm{vol}(v_1, v_2) = -\,\mathrm{vol}(v_2, v_1).$$

We can now calculate the volume of a parallelogram, using these formal properties of vol and the fact that the volume of a square is

$$\mathrm{vol}(e_1, e_2) = 1$$

for e_1, e_2 the standard basis of $\mathbb{R}^2$.

Proposition 1.5. *Let* $v_1, v_2 \in \mathbb{R}^2$ *be the column vectors of a matrix* A. *Then*

$$\mathrm{vol}(v_1, v_2) = \det(A).$$

Proof. Write

$$v_1 = \begin{pmatrix} a \\ c \end{pmatrix} = ae_1 + ce_2, \quad v_2 = \begin{pmatrix} b \\ d \end{pmatrix} = be_1 + de_2.$$

Using bi-linearity to expand, we find

$$\begin{aligned} \mathrm{vol}(v_1, v_2) &= a\,\mathrm{vol}(e_1, v_2) + c\,\mathrm{vol}(e_2, v_2) \\ &= ac\,\mathrm{vol}(e_1, e_1) + ad\,\mathrm{vol}(e_1, e_2) + cb\,\mathrm{vol}(e_2, e_1) + cd\,\mathrm{vol}(e_2, e_2) \\ &= ad - bc \\ &= \det(A). \end{aligned}$$

□

Although this interpretation as an area only works for $F = \mathbb{R}$, we can generalize the definition of vol to arbitrary F – although it seems reasonable now to rename it as det.

Namely, we see that there is a *unique bi-linear functional*

$$\det\colon F^2 \times F^2 \to F, \quad (v_1, v_2) \mapsto \det(v_1, v_2)$$

such that $\det(v, v) = 0$ for all $v \in F^2$, and with $\det(e_1, e_2) = 1$ for the standard basis. In fact, the calculation above shows that $\det(v_1, v_2) = \det(A) = ad - bc$.

Remark 1.6. If $\phi\colon V \times V \to F^2$ is a bilinear functional on a vector space V, then

$$\phi(v, v) = 0 \quad \text{for all} v \in V \quad \Rightarrow \quad \phi(v_1, v_2) = -\phi(v_2, v_1) \;\text{ for all }\; v_1, v_2 \in V.$$

Is this an equivalence? Only if the characteristic of the field is $\neq 2$. In fact we have

$$\phi(v_1, v_2) = -\phi(v_2, v_1) \;\text{ for all }\; v_1, v_2 \in V \quad \Rightarrow \quad 2\phi(v, v) = 0 \quad \text{for all} v \in V$$

(this follows by putting $v_1, v_2 = v$). Thus, if $2 \neq 0$ in F we can divide by 2, and we recover $\phi(v, v) = 0$. On the other hand, if $2 = 0$ in F, this conclusion is wrong in general. E.g., the bilinear functional

$$\phi\Big(\begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} \Big) = ab + cd$$

(dot product) on F^2 is symmetric. If $2 = 0$ in F, then $1 = -1$, and so symmetric forms are also skew-symmetric. But it does not satisfy $\phi(v, v) = 0$ for all v.

1.3. **Generalization to higher dimensions.** In $\mathbb{R}^n$, we consider the signed volume of the parallelepiped spanned by $v_1, \ldots, v_n$, denoted $\mathrm{vol}(v_1, \ldots, v_n)$. If the v_i are the standard basis vectors, we get the volume of the unit cube: $\mathrm{vol}(e_1, \ldots, e_n) = 1$. As above, we find that this is linear in each argument. For general fields, we use these properties to define a 'volume function'. Generalizing to arbitrary fields, we have

Theorem 1.7. *There exists a unique multi-linear functional*

$$\det\colon F^n \times \cdots \times F^n \to F$$

with the property that $\det(v_1, \ldots, v_n) = 0$ *whenever two of the* v_i*'s coincide, and with*

$$\det(e_1, \ldots, e_n) = 1,$$

for the standard ordered basis $e_1, \ldots, e_n$ *of* F^n*.*

Here, *multi-linear* means that det is linear in each argument, keeping the others fixed. E.g.,

$$\det(v_1,\ldots,v_{i-1},v_i+v_i',v_{i+1},\ldots)=\det(v_1,\ldots,v_{i-1},v_i',v_{i+1},\ldots)+\det(v_1,\ldots,v_{i-1},v_i',v_{i+1},\ldots).$$

and

$$\det(v_1,\ldots,v_{i-1},av_i,v_{i+1},\ldots)=a\det(v_1,\ldots,v_{i-1},v_i,v_{i+1},\ldots).$$

Before proving the theorem, a few facts about *permutations*.

Definition 1.8. A *permutation* of $\{1,\ldots,n\}$ is an invertible map, σ from this set to itself. The permutation is called *even* (resp. odd) if the number of pairs (i_1,i_2) such that $i_1<i_2$ but $\sigma(i_1)>\sigma(i_2)$ is even (resp. odd) . One writes $\text{sign}(\sigma)=1$ resp. -1 depending on whether the permutation is even or odd.

Example 1.9. Here $n=4$. The permutation

$$\sigma(1)=4,\ \sigma(2)=3,\ \sigma(3)=1,\ \sigma(4)=2,$$

depicted as

$$(4,3,1,2),$$

is odd; $\text{sign}(\sigma)=-1$, because there are five pairs of indices in wrong order,

$$(4,3),(4,1),(4,2),(3,1),(3,2).$$

Note that if one modifies a permutation by interchanging two adjacent elements, then the parity of σ changes. Namely, the ordering of that pair changes from right to wring or the other way; whereas all other orderings are preserved.

Example 1.10. In the example above, the permutation σ' written as $(4,1,3,2)$ (obtained by switching 1 and 3 in σ) is even: $\text{sign}(\sigma')=1$.

By induction, we conclude that for any permutation σ, we have that $\text{sign}(\sigma)=(-1)^N$ if one can put the elements back into their original order by N transpositions of adjacent elements.

Example 1.11.

$$(4,3,1,2)\to(4,1,3,2)\to(1,4,3,2)\to(1,4,2,3)\to(1,2,4,3)\to(1,2,3,4).$$

Here $N=5$, so we recover that $\text{sign}(\sigma)=-1$.

Actually, one can speed up calculations a bit using the following

Exercise. Show that if σ' is obtained from σ by interchanging two elements (not necessariy adjacent), then σ',σ have opposite parity.

Example 1.12.

$$(4,3,1,2)\to(1,3,4,2)\to(1,2,4,3)\to(1,2,3,4)$$

Here $N=3$ so $\text{sign}(\sigma)=-1$.

Let us return to the proof of the theorem.

Proof. We start with the uniqueness proof (assuming existence.) Any multi-linear functional is uniquely determined by its values on n-tuples of basis vectors, since the general formula then follows by multi-linearity. Thus, we need to specify

$$\det(e_{i_1}, \ldots, e_{i_n})$$

for arbitrary $i_1, \ldots, i_n \in \{1, \ldots, n\}$. By assumption, this has to be zero if two of the indices coincide. So, the only case one gets something non-zero is if

$$i_1 = \sigma(1),\ i_1 = \sigma(2),\ \ldots,\ i_n = \sigma(n)$$

for some *permutation* of the indices. In that case, we can put $e_{\sigma(1)}, \ldots, e_{\sigma(n)}$ into the right order by a finite number of interchanges ('transposition') of indices. As in the case $n = 2$, we see that the interchange of any two arguments of det gives a minus sign. Thus we must have

$$\det(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = \operatorname{sign}(\sigma) \det(e_1, \ldots, e_n) = \operatorname{sign}(\sigma).$$

Consider now general vectors $v_j \in F^n$, expressed in terms of the basis as

$$v_j = \sum_i A_{ij} e_i.$$

By multi-linearity,

$$\begin{aligned} \det(v_1, \ldots, v_n) &= \sum_{i_1 \cdots i_n} \det(A_{i_1,1} e_{i_1}, \cdots, A_{i_n,n} e_{i_n}) \\ &= \sum_{i_1 \cdots i_n} A_{i_1,1} \cdots A_{i_n,n} \det(e_{i_1}, \ldots, e_{i_n}) \end{aligned}$$

As we just mentioned, the summand are zero unless$i_1, \ldots, i_n$ are a permutation of $1, \ldots, n$. We thus obtain

$$\det(v_1, \ldots, v_n) = \sum_\sigma \operatorname{sign}(\sigma)\ A_{\sigma(1),1} \cdots A_{\sigma(n),n}.$$

This explicit formula shows that det is uniquely determined by its properties.

For existence, we use this formula as a definition of a multi-linear functional. Clearly, with this definition $\det(e_1, \ldots, e_n) = 1$, because in this case $A_{ij} = \delta_{ij}$ and only the trivial permutation $\sigma = \mathrm{id}$ contributes.

We have to show that $\det(v_1, \ldots, v_n)$ vanishes whenever $v_r = v_s$ for some $r < s$. In this case we have that $A_{ir} = A_{is}$ for all $i = 1, \ldots, n$. If σ is any permutation, there is a unique permutation $\sigma' \neq \sigma$ such that $\sigma(i) = \sigma'(i)$ for all $i \neq r, s$. In fact,

$$\sigma'(i) = \begin{cases} \sigma(i) & i \neq r, s \\ \sigma(s) & i = r \\ \sigma(r) & i = s. \end{cases}$$

Since

$$A_{\sigma(r),r} A_{\sigma(s),s} = A_{\sigma(r),s} A_{\sigma(s),r} = A_{\sigma'(r),r} A_{\sigma'(s),s}$$

we get

$$A_{\sigma(1),1} \cdots A_{\sigma(n),n} = A_{\sigma'(1),1} \cdots A_{\sigma'(n),n}.$$

Note that the two permutations σ, σ' have opposite sign, since one is obtained from the other by interchanging two elements: $\operatorname{sign}(\sigma) = -\operatorname{sign}(\sigma')$. It follows that the corresponding terms in the sum cancel. We conclude $\det(v_1, \ldots, v_n) = 0$. $\square$

After all this hard work, we can finally define:

Definition 1.13. The *determinant* of a square matrix $A \in M_{n\times n}(F)$ is defined as

$$\det(A) = \det(v_1, \ldots, v_n),$$

where $v_1, \ldots, v_n$ are the columns of A.

The proof above gave us a formula for the determinant:

$$\boxed{\det(A) = \sum_\sigma \operatorname{sign}(\sigma)\ A_{\sigma(1)1} \cdots A_{\sigma(n)n}.}$$

If $n = 2$ we recover the formula $\det(A) = A_{11}A_{22} - A_{21}A_{12}$.

Remark 1.14. There are a number of methods of computing determinants. The (complicated) formula is not very efficient in practice (except for $n \le 2$), since the number of terms of this expression is $n!$ (the number of permutations). E.g. for 5×5 matrices we already get 120 terms!

Theorem 1.15 (Properties of the determinant)**.** *Let* $A, B \in M_{n\times n}(F)$*.*

(a) *The determinant* $\det(A)$ *vanishes if and only if the columns of* A *are linearly dependent.*
(b) *If* A' *is obtained from* A *by interchange of two columns, the* $\det(A') = -\det(A)$*.*
(c) *If* A' *is obtained from* A *by taking the* c*-th multiple of one column, the* $\det(A') = c\det(A)$*.*
(d) *If* A' *s obtained from* A *by adding a scalar multiple of one column to another column, then* $\det(A') = \det(A)$*.*
(e) $\det(A^t) = \det(A)$*; hence the above statements also hold for columns replaced with rows.*
(f) $\det(AB) = \det(A)\det(B)$*. In particular,* $\det(A^{-1}) = \det(A)^{-1}$*.*

Proof. By construction, the determinant function $A \mapsto \det(A)$ is linear in the columns of A, and vanishes whenever two columns coincide. This already implies (c), as well as (d). As in the case $n = 2$, the fact that $\det(A)$ vanishes whenever two of the columns are equal, implies that it changes sign under exchange of two columns, i.e. (b).

Using column operations, we may bring A into reduced column echelon form A' (which amounts to using row operations on A^t to bring A^t to reduced row echelon form). By (b),(c),(d) this changes the determinant by a non-zero scalar. If $\operatorname{rank}(A) < n$, it then follows that some column of A' is zero, hence $\det(A') = 0$ by linearity. We then conclude $\det(A) = 0$. If $\operatorname{rank}(A) = n$, then A' is the identity matrix, hence $\det(A') = 1$. We conclude $\det(A) \neq 0$. This proves (a).

Property (e) follows from the explicit 'complicated formula', using the fact that $\operatorname{sign}(\sigma^{-1}) = -\operatorname{sign}(\sigma)$, or using (f) and the fact that every matrix can be written as a product of elementary matrices. (For elementary matrices, the property is obvious.)

For property (f), we argue as follows. If A is not invertible, then AB is also not invertible, and both sides are zero. Hence we may assume that A is invertible. The multilinear functional

$$\phi(w_1, \ldots, w_n) = \frac{\det(Aw_1, \ldots, Aw_n)}{\det(A)}$$

vanishes if any two of the w_i coincide, and $\phi(e_1, \ldots, e_n) = 1$ (since Ae_i are the columns of A). Hence $\phi = \det$. Now take $w_i = Be_i$, the columns of B. Then

$$\det(w_1, \ldots, w_n) = \det(B),$$

$$\det(Aw_1, \ldots Aw_n) = \det(AB(e_1), \ldots, AB(e_n)) = \det(AB).$$

We conclude $\det(B) = \det(AB)/\det(A)$. □

Part (a) of this theorem has a very important consequence: A square matrix $A \in M_{n\times n}$ is invertible if and only if $\det(A) \neq 0$. In particular, in this case the equation $Ax = b$ has a unique solution for all $b \in F^n$. In fact, there is a simple formula expressing the solution in terms of determinants.

Theorem 1.16 (Cramer's rule)**.** *Let $A \in M_{n\times n}$ be an invertible matrix, with columns $v_1, \ldots, v_n$. Then the unique solution $x = (x_1, \ldots, x_n)^t$ to the equation $Ax = b$ is given by the formula*

$$x_i = \frac{1}{\det A}\ \det(v_1, \ldots, v_{i-1}, b, v_{i+1}, \ldots, v_n).$$

(Thus, for each i one takes the determinant of the matrix obtained by replacing the i-th column v_i with b, and divides by $\det(A)$*.)*

Proof. The unique solution is, of course, $x = A^{-1}b$. By definition of matrix multiplication,

$$b = Ax = x_1v_1 + \ldots + x_nv_n.$$

Thus, expending by linearity in the ith column,

$$\det(v_1, \ldots, v_{i-1}, b, v_{i+1}, \ldots, v_n) = \sum_{r=1}^{n} x_r \det(v_1, \ldots, v_{i-1}, v_r, v_{i+1}, \ldots, v_n).$$

But $\det(v_1, \ldots, v_{i-1}, v_r, v_{i+1}, \ldots, v_n) = 0$ unless $r = i$, in which case it is $\det(A)$. This shows

$$\det(v_1, \ldots, v_{i-1}, b, v_{i+1}, \ldots, v_n) = x_i \det(A).$$

□

For invertible matrices, this is a rather useful formula – provided we learn how to calculate determinants.

Example 1.17. The solution of the equation $Ax = b$, for $A \in M_{3\times 3}(\mathbb{R})$ given as

$$A = \begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 4 \\ -3 & -2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 7 \\ -1 \end{pmatrix}$$

is

$$x_1 = \frac{\det\begin{pmatrix} 1 & 0 & -1 \\ 7 & 2 & 4 \\ -1 & -2 & 1 \end{pmatrix}}{\det\begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 4 \\ -3 & -2 & 1 \end{pmatrix}}, \quad x_2 = \frac{\det\begin{pmatrix} 3 & 1 & -1 \\ 0 & 7 & 4 \\ -3 & -1 & 1 \end{pmatrix}}{\det\begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 4 \\ -3 & -2 & 1 \end{pmatrix}}, \quad x_3 = \frac{\det\begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 7 \\ -3 & -2 & -1 \end{pmatrix}}{\det\begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 4 \\ -3 & -2 & 1 \end{pmatrix}}$$

We'll see below how to efficiently calculate the determinants.

Note that Cramer's rule also gives a formula for the inverse matrix A^{-1}. Let $(v_1, \ldots, v_n)$ be the columns of A, and $w_1, \ldots, w_n$ the columns of A^{-1}. Thus $w_j = A^{-1}e_j$, i.e., w_j is the solution to $Ax = e_j$, and the matrix entry $(A^{-1})_{ij}$ is the i-th component of this solution. Thus, by Cramer's rule

$$(A^{-1})_{ij} = \frac{1}{\det(A)} \det(v_1, \ldots, v_{i-1}, e_j, v_{i+1}, \ldots, v_n).$$

To calculate $\det(v_1, \ldots, v_{i-1}, e_j, v_{i+1}, \ldots, v_n)$, note that we can use e_j to clear out all entries in the j-th row.