# Notes on linear algebra

Danny Nygård Hansen

6th February 2022

## 1 • Linear equations and matrices

### *1.1. Linear equations*

Throughout we let $F$ denote an arbitrary field and $R$ a commutative ring. Let $m$ and $n$ be positive integers. A *linear equation in n unknowns* is an equation on the form

$$l\colon a_1x_1 + \cdots + a_nx_n = b,$$

where $a_1,\dots,a_n,b \in F$. A *solution* to $l$ is an element $v = (v_1,\dots,v_n) \in F^n$ such that

$$a_1v_1 + \cdots + a_nv_n = b.$$

A *system of linear equations in n unknowns* is a tuple $L = (l_1,\dots,l_m)$, where each $l_i$ is a linear equation in $n$ unknowns. An element $v \in F^n$ is a *solution* to $L$ if it is a solution to each linear equation $l_1,\dots,l_m$.

Let $L$ and $L'$ be systems of linear equations in $n$ unknowns. We say that $L$ and $L'$ are *solution equivalent* if they have the same solutions. Furthermore, we say that they are *combination equivalent* if each equation in $L'$ is a linear combination of the equations in $L$, and vice versa. Clearly, if $L$ and $L'$ are combination equivalent they are also solution equivalent, but the converse does not hold.

### *1.2. Matrices*

It is well-known that a system of linear equations is equivalent to a matrix equation on the form $Ax = b$, where $A \in \mathcal{M}_{m,n}(F)$, $x \in F^n$ and $b \in F^m$. Recall the *elementary row operations* on $A$:

(1) multiplication of one row of $A$ by a nonzero scalar,

(2) addition to one row of $A$ a scalar multiple of another (different) row, and

(3) interchange of two rows of $A$.

If $e$ is an elementary row operation, we write $e(A)$ for the matrix obtained when applying $e$ to $A$. Clearly each elementary row operation $e$ has an 'inverse', i.e. an elementary row operation $e'$ such that $e'(e(A)) = e(e'(A)) = A$. Two matrices $A, B \in \mathcal{M}_{m,n}(F)$ are called *row-equivalent* if $A$ is obtained by applying a finite sequence of elementary row operations to $B$ (and vice versa, though this need not be assumed since each elementary row operation has an inverse).

Clearly, if $A, B \in \mathcal{M}_{m,n}(F)$ are row-equivalent, then the systems of equations $Ax = 0$ and $Bx = 0$ are combination equivalent, hence have the same solutions.

**Definition 1.1**

A matrix $H \in \mathcal{M}_{m,n}(F)$ is called *row-reduced* if

(i) the first nonzero entry of each nonzero row in $H$ is 1, and

(ii) each column of $H$ containing the leading nonzero entry of some row has all its other entries equal 0.

If $H$ is row-reduced, it is called a *row-reduced echelon matrix* if it also has the following properties:

(iii) Every row of $H$ only containing zeroes occur below every row which has a nonzero entry, and

(iv) if rows $1, \dots, r$ are the nonzero rows of $H$, and if the leading nonzero entry of row $i$ occurs in column $k_i$, then $k_1 < \cdots < k_r$.

An *elementary matrix* is a matrix obtained by applying a single elementary row operation to the identity matrix $I$. It is easy to show that if $e$ is an elementary row operation and $E = e(I) \in \mathcal{M}_m(F)$, then $e(A) = EA$ for $A \in \mathcal{M}_{m,n}(F)$. If $B \in \mathcal{M}_{m,n}(F)$, then $A$ and $B$ are row-equivalent if and only if $A = PB$, where $P \in \mathcal{M}_m(F)$ is a product of elementary matrices.

**Proposition 1.2**

*Every matrix in $\mathcal{M}_{m,n}(F)$ is row-equivalent to a unique row-reduced echelon matrix.*

Proof. The usual Gauss–Jordan elimination algorithm proves existence. If $H, K \in \mathcal{M}_{m,n}(R)$ are row-equivalent row-reduced echelon matrices, we claim that $H = K$. We prove this by induction in $n$. If $n = 1$ then this is obvious, so assume that $n > 1$. Let $H_1$ and $K_1$ be the matrices obtained by deleting the $n$th column in $H$ and $K$ respectively. Then $H_1$ and $K_1$ are also row-equivalent[1] and row-reduced echelon matrices, so by induction $H_1 = K_1$. Thus if $H$ and $K$ differ, they must differ in the $n$th column.

Let $H_2$ be the matrix obtained by deleting columns in $H$, only keeping those columns containing pivots, as well as keeping the $n$th column. Define $K_2$ similarly. Thus we have deleted the same columns in $H$ and $K$, so $H_2$ and $K_2$ are also row-equivalent. Say that the number of columns in $H_2$ and $K_2$ is $r+1$, and write the matrices on the form

$$H_2 = \begin{pmatrix} I_r & h \\ 0 & h' \end{pmatrix} \quad \text{and} \quad K_2 = \begin{pmatrix} I_r & k \\ 0 & k' \end{pmatrix},$$

where $h, k \in F^r$ and $h', k' \in F^{m-r}$ are column vectors. Since $H_2$ and $K_2$ are row-equivalent, the systems $H_2 x = 0$ and $K_2 x = 0$ are solution equivalent. If $h' = 0$, then $H_2 x = 0$ has the solution $(-h, 1)$. But this is also a solution to $K_2 x = 0$, so $h = k$ and $k' = 0$. If $h' \neq 0$, then $H_2 x = 0$ only has the trivial solution. But then $K_2 x = 0$ also only has the trivial solution, and hence $k' \neq 0$. But that must be because both $H_2$ and $K_2$ has a pivot in the $n$th column, so also in this case $H_2 = K_2$. □

## 1.3. *Invertible matrices*

Notice that elementary matrices are invertible, since elementary row operations are invertible.

**LEMMA 1.3**

*If $A \in \mathcal{M}_n(F)$, then the following are equivalent:*

(i) *$A$ is invertible,*

(ii) *$A$ is row-equivalent to $I_n$,*

(iii) *$A$ is a product of elementary matrices, and*

(iv) *the system $Ax = 0$ has only the trivial solution $x = 0$.*

PROOF. *(i) $\Leftrightarrow$ (ii)*: Let $H \in \mathcal{M}_n(F)$ be a row-reduced echelon matrix that is row-equivalent to $A$. Then $R = PA$, where $P \in \mathcal{M}_n(F)$ is a product of elementary matrices. Then $A = P^{-1}H$, so $A$ is invertible if and only if $H$ is. But the only invertible row-reduced echelon matrix is the identity matrix, so (i) and (ii) are equivalent.

[1] It should be obvious that deleting columns preserves row-equivalence, but we give a more precise argument: If $P \in \mathcal{M}_m(F)$ is a product of elementary matrices and $a_1, \ldots, a_n \in F^m$ are the columns in $A$, then the columns in $PA$ are $Pa_1, \ldots, Pa_m$. Thus elementary row operations are applied to each column independently of the other columns.

*(i) ⇔ (iii)*: Clearly (iii) implies (i), and the above shows that (i) implies that $A = P^{-1}$.

*(ii) ⇔ (iv)*: If $A$ and $I_n$ are row-equivalent, then the systems $Ax = 0$ and $I_n x = 0$ have the same solutions. Conversely, assume that $Ax = 0$ only has the trivial solution. If $H \in \mathcal{M}_{m,n}(F)$ is a row-reduced echelon matrix that is row-equivalent to $A$, then $Hx = 0$ has no nontrivial solution. Thus if $r$ is the number of nonzero rows in $H$, then $r \geq n$. But then $r = n$, so $H$ must be the identity matrix. □

**PROPOSITION 1.4**

*Let $A \in \mathcal{M}_n(F)$. Then the following are equivalent:*

(i) *$A$ is invertible,*

(ii) *$A$ has a left inverse, and*

(iii) *$A$ has a right inverse.*

PROOF. If $A$ has a left inverse, then $Ax = 0$ has no nontrivial solution, so $A$ is invertible. If $A$ has a right inverse $B \in \mathcal{M}_n(F)$, i.e. $AB = I$, then $B$ has a left inverse and is thus invertible. But then $A$ is the inverse of $B$ and hence is itself invertible. □

# 2 • Determinants

## *2.1. Existence of determinants*

If $M_1, \dots, M_n, N$ are modules over a commutative ring $R$, a map

$$\varphi \colon M_1 \times \cdots \times M_n \to N$$

is called *$n$-linear* if the maps $m_i \mapsto \varphi(m_1, \dots, m_n)$ are linear for all $m_i \in M_i$. Since $\mathcal{M}_{m,n}(R) \cong (R^m)^n$, a map $\varphi \colon \mathcal{M}_{m,n}(R) \to N$ that is linear in each row is also called $n$-linear.

In the case $M_1 = \cdots = M_n$, we call $\varphi$ *alternating* if $\varphi(m_1, \dots, m_n) = 0$ whenever $m_i = m_j$ for some $i \neq j$. Furthermore, $\varphi$ is called *skew-symmetric* if

$$\begin{aligned}\varphi(m_1, \dots, m_{i-1}, m_i, m_{i+1}, \dots, m_{j-1}, m_j, m_{j+1}, \dots, m_n)& \\ = -\varphi(m_1, \dots, m_{i-1}, m_j, m_{i+1}, \dots, m_{j-1}, m_i, m_{j+1}, \dots, m_n)&\end{aligned}$$

for all $i < j$.

LEMMA 2.1

*Let $M$ and $N$ be $R$-modules, and let $\varphi\colon M^n \to N$ be an $n$-linear map.*

(i) *If $\varphi$ is alternating, then $\varphi$ is skew-symmetric.*

(ii) *If $\varphi(m_1,\dots,m_n) = 0$ whenever $m_i = m_{i+1}$ for some $i = 1,\dots,n-1$, then $\varphi$ is alternating.*

PROOF. *(i)*: Consider $m_1,\dots,m_n \in M$, and let $1 \leq i < j \leq n$. Define a map $\psi\colon M \times M \to N$ by

$$\psi(a,b) = \varphi(m_1,\dots,m_{i-1},a,m_{i+1},\dots,m_{j-1},b,m_{j+1},\dots,m_n),$$

and notice that it suffices to show that $\psi(m_i,m_j) = -\psi(m_j,m_i)$. But $\psi$ is 2-linear and alternating, so for $a,b \in M$ we have

$$\psi(a+b,a+b) = \psi(a,a)+\psi(a,b)+\psi(b,a)+\psi(b,b) = \psi(a,b)+\psi(b,a).$$

Thus $\psi(m_i,m_j) = -\psi(m_j,m_i)$, so $\varphi$ is skew-symmetric as claimed.

*(ii)*: The argument above shows that, in particular, if $A,B \in M^n$, and $B$ is obtained from $A$ by interchanging two adjacent elements, then $\varphi(B) = -\varphi(A)$. Assuming now that $B$ is obtained from $A$ by interchanging the $i$th and $j$th elements in $A$, with $i < j$, we claim that we may obtain $B$ by successively interchanging adjacent elements of $A$. Writing $A = (m_1,\dots,m_n)$, we first perform $j-i$ such interchanges and arrive that the tuple

$$(m_1,\dots,m_{i-1},m_{i+1},\dots,m_{j-1},m_j,m_i,m_{j+1},\dots,m_n),$$

moving $m_i$ to the right $j-i$ places. Next we perform another $j-i-1$ interchanges, moving $m_j$ to the left until we reach

$$B = (m_1,\dots,m_{i-1},m_j,m_{i+1},\dots,m_{j-1},m_i,m_{j+1},\dots,m_n).$$

Since each interchange results in a sign change, we have

$$\varphi(B) = (-1)^{2(j-i)-1}\varphi(A) = -\varphi(A).$$

If $m_i = m_j$ for $i < j$, then we claim that $\varphi(A) = 0$. For let $B$ be obtained from $A$ by interchanging $m_{i+1}$ and $m_j$. Then $\varphi(B) = 0$, so $\varphi(A) = -\varphi(B) = 0$ by the above argument, and hence $\varphi$ is alternating as claimed. □

DEFINITION 2.2

If $n$ be a positive integer, a *determinant function* is a map $\varphi\colon \mathcal{M}_n(R) \to R$ that is $n$-linear, alternating, and which satisfies $\varphi(I_n) = 1$.

If $A \in \mathcal{M}_n(R)$ with $n > 1$ and $1 \leq i, j \leq n$, denote by $M(A)_{i,j}$ the matrix in $\mathcal{M}_{n-1}(R)$ obtained by removing the the $i$th row and the $j$th column of $A$. This is called the $(i,j)$-*th minor* of $A$. If $\varphi\colon \mathcal{M}_{n-1}(R) \to R$ is an $(n-1)$-linear function and $A \in \mathcal{M}_n(R)$, then we write $\varphi_{i,j}(A) = \varphi(M(A)_{i,j})$. Then $\varphi_{i,j}\colon \mathcal{M}_n(R) \to R$ is clearly linear in all rows except row $i$, and is independent of row $i$.

THEOREM 2.3

*Let $n > 1$, and let $\varphi\colon \mathcal{M}_{n-1}(R) \to R$ be alternating and $(n-1)$-linear. For $j = 1, \dots, n$ define a map $\psi_j\colon \mathcal{M}_n(R) \to R$ by*

$$\psi_j(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \varphi_{i,j}(A),$$

*for $A = (a_{ij}) \in \mathcal{M}_n(R)$. Then $\psi_j$ is alternating and $n$-linear. If $\varphi$ is a determinant function, then so is $\psi_j$.*

PROOF. Let $A = (a_{ij}) \in \mathcal{M}_n(R)$. Then $A \mapsto a_{ij}$ is independent of all rows except row $i$, and $\varphi_{i,j}$ is linear in all rows except row $i$. Thus $A \mapsto a_{ij}\varphi_{i,j}(A)$ is linear in all rows except row $i$. Conversely, $A \mapsto a_{ij}$ is linear in row $i$, and $\varphi_{i,j}$ is independent of row $i$, so $A \mapsto a_{ij}\varphi_{i,j}(A)$ is also linear in row $i$. Since $\psi_j$ is a linear combination of $n$-linear maps, is it itself $n$-linear.

Now assume that $A$ has two equal adjacent rows, say $a_k, a_{k+1} \in R^n$. If $i \neq k$ and $i \neq k+1$, then $M(A)_{i,j}$ has two equal rows, so $\varphi_{i,j}(A) = 0$. Thus

$$\psi_j(A) = (-1)^{k+j} a_{kj} \varphi_{k,j}(A) + (-1)^{k+1+j} a_{(k+1)j} \varphi_{k+1,j}(A).$$

Since $a_k = a_{k+1}$ we also have $a_{kj} = a_{(k+1)j}$ and $M(A)_{k,j} = M(A)_{k+1,j}$. Thus $\psi_j(A) = 0$, so Lemma 2.1(ii) implies that $\psi_j$ is alternating.

Finally suppose that $\varphi$ is a determinant function. Then $M(I_n)_{j,j} = I_{n-1}$ and we have

$$\psi_j(I_n) = (-1)^{j+j} \varphi_{j,j}(I_n) = \varphi(I_{n-1}) = 1,$$

so $\psi_j$ is also a determinant function. □

COROLLARY 2.4

*For every positive integer $n$, there exists a determinant function $\mathcal{M}_n(R) \to R$.*

PROOF. The identity map on $\mathcal{M}_1(R) \cong R$ is a determinant function for $n = 1$, and Theorem 2.3 allows us to recursively construct a determinant for each $n > 1$. □

## 2.2. Uniqueness of determinants

**Theorem 2.5**

*Let $n$ be a positive integer. There is precisely one determinant function on $\mathcal{M}_n(R)$, namely the function* $\det\colon \mathcal{M}_n(R) \to R$ *given by*

$$\det A = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

*for $A = (a_{ij}) \in \mathcal{M}_n(R)$. If $\varphi\colon \mathcal{M}_n(R) \to R$ is any alternating $n$-linear function, then*

$$\varphi(A) = (\det A)\varphi(I_n).$$

We use the notation det for the unique determinant on $\mathcal{M}_n(R)$ for all $n$.

Proof. Let $e_1, \dots, e_n$ denote the rows of $I_n$, and denote the rows of a matrix $A = (a_{ij}) \in \mathcal{M}_n(R)$ by $a_1, \dots, a_n$. Then $a_i = \sum_{j=1}^n a_{ij} e_j$, so

$$\varphi(A) = \sum_{k_1, \dots, k_n} a_{1k_1} \cdots a_{nk_n} \varphi(e_{k_1}, \dots, e_{k_n}),$$

where the sum is taken over all $k_i = 1, \dots, n$. Since $\varphi$ is alternating we have $\varphi(e_{k_1}, \dots, e_{k_n}) = 0$ if two of the indices $k_1, \dots, k_n$ are equal. Thus it suffices to sum over those sequences $(k_1, \dots, k_n)$ that are permutations of $(1, \dots, n)$, and so

$$\varphi(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \varphi(e_{\sigma(1)}, \dots, e_{\sigma(n)}).$$

Next notice that, since $\varphi$ is also skew-symmetric by Lemma 2.1(i), we have $\varphi(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = (-1)^m \varphi(e_1, \dots, e_n)$, where $m$ is the number of transpositions of $(1, \dots, n)$ it takes to obtain the permutation $(\sigma(1), \dots, \sigma(n))$. But then $(-1)^m$ is just the sign of $\sigma$, so

$$\varphi(A) = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \varphi(I_n).$$

Finally, if $\varphi$ is a determinant function, then $\varphi(I_n) = 1$, so we must have $\varphi = \det$. The rest of the theorem follows directly from this. □

## 2.3. Properties of determinants

**Theorem 2.6**

*Let $A, B \in \mathcal{M}_n(R)$. Then*

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

*In particular,* $\det\colon \mathrm{GL}_n(R) \to R^*$ *is a group homomorphism.*

Proof. The map $\varphi\colon \mathcal{M}_n(R) \to R$ given by $\varphi(A) = \det AB$ is clearly $n$-linear and alternating. Hence $\varphi(A) = (\det A)\varphi(I)$, and $\varphi(I) = \det B$.

Furthermore, if $A$ is invertible, then $1 = \det I = (\det A)(\det A^{-1})$. Thus $\det A \in R^*$, so det is a group homomorphism as claimed. □

Proposition 2.7

*Let* $B_{11},\dots,B_{nn}$ *be square matrices with entries in* $R$ *and consider the block matrix*

$$A = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ 0 & B_{22} & \ddots & B_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & B_{nn} \end{pmatrix},$$

*where the remaining* $B_{ij}$ *are matrices of appropriate dimensions. Then* $\det A = \prod_{i=1}^{n} \det B_{ii}$.

Proof. By induction it suffices to consider the case where $A$ has the block form

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}.$$

Say that $B \in \mathcal{M}_r(R)$ and $D \in \mathcal{M}_s(R)$, and put $\varphi(B,C,D) = \det A$. Then $D \mapsto \varphi(B,C,D)$ is clearly $s$-linear and alternating, so Theorem 2.3 implies that

$$\varphi(B,C,D) = (\det D)\varphi(B,C,I_s).$$

By subtracting multiples of the rows of $I_s$ from $C$ we obtain $\varphi(B,C,I_s) = \varphi(B,0,I_s)$. Next, $B \mapsto \varphi(B,0,I_s)$ is also $r$-linear and alternating, so

$$\varphi(B,0,I_s) = (\det B)\varphi(I_r,0,I_s).$$

But $\varphi(I_r,0,I_s) = 1$, so summarising we have

$$\varphi(B,C,D) = (\det D)\varphi(B,C,I_s) = (\det D)\varphi(B,0,I_s) = (\det D)(\det B),$$

as desired. □

Proposition 2.8

*Let* $A \in \mathcal{M}_n(R)$. *Then* $\det A = \det A^\top$.

PROOF. Writing $A = (a_{ij})$, first notice that

$$\det A^\top = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma^{-1}) a_{\sigma(1)1} \cdots a_{\sigma(n)n},$$

since $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma^{-1}$. Next notice that, if $j = \sigma(i)$, then $a_{\sigma(i)i} = a_{j\sigma^{-1}(j)}$. Since $R$ is commutative, it follows that

$$\det A^\top = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)},$$

and since $\sigma \mapsto \sigma^{-1}$ is a bijection on $S_n$, it follows that $\det A^\top = \det A$ as desired. □

Let $A \in \mathcal{M}_n(R)$. For $1 \le i, j \le n$, the $(i,j)$*-th cofactor* of $A$ is the number $A_{i,j} = (-1)^{i+j} \det M(A)_{i,j}$, where we recall that $M(A)_{i,j}$ is the $(i,j)$-th minor of $A$. The *adjoint matrix* of $A$ is the matrix $\operatorname{adj} A \in \mathcal{M}_n(R)$ whose $(i,j)$-th entry is the cofactor $A_{j,i}$. Note that

$$(A^\top)_{i,j} = (-1)^{i+j} \det M(A^\top)_{i,j} = (-1)^{j+i} \det M(A)_{j,i} = A_{j,i},$$

so $\operatorname{adj} A^\top = (\operatorname{adj} A)^\top$. We have the following:

PROPOSITION 2.9

*Let $A \in \mathcal{M}_n(R)$. Then*

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

PROOF. Writing $A = (a_{ij})$ and fixing some $j \in \{1, \dots, n\}$, Theorem 2.3 implies that

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det M(A)_{i,j} = \sum_{i=1}^n a_{ij} A_{i,j},$$

which is just the $(j,j)$-th entry in the product $(\operatorname{adj} A)A$.

Next we claim that if $k \ne j$, then $\sum_{i=1}^n a_{ik} A_{i,j} = 0$. Let $B = (b_{ij}) \in \mathcal{M}_n(R)$ be the matrix obtained from $A$ by replacing the $j$th column of $A$ by its $k$th column. Then $B$ has two equal columns, so $\det B = 0$. Also, $b_{ij} = a_{ik}$ and $M(B)_{i,j} = M(A)_{i,j}$, so it follows that

$$\begin{aligned} 0 = \det B &= \sum_{i=1}^n (-1)^{i+j} b_{ij} \det M(B)_{i,j} \\ &= \sum_{i=1}^n (-1)^{i+j} a_{ik} \det M(A)_{i,j} = \sum_{i=1}^n a_{ik} A_{i,j}. \end{aligned}$$

That is, the $(j, k)$-th entry of the product $(\operatorname{adj} A)A$ is zero, so the off-diagonal entries of $(\operatorname{adj} A)A$ are zero. In total we thus have $(\operatorname{adj} A)A = (\det A)I$.

Finally we prove the equality $A(\operatorname{adj} A) = (\det A)I$, Applying the first equality to $A^\top$ yields

$$(\operatorname{adj} A^\top)A^\top = (\det A^\top)I = (\det A)I,$$

and transposing we get

$$A(\operatorname{adj} A) = A(\operatorname{adj} A^\top)^\top = (\det A)I$$

as desired. □

**COROLLARY 2.10**

*Let $A \in \mathcal{M}_n(R)$. Then $A$ is a unit in $\mathcal{M}_n(R)$ if and only if $\det A$ is a unit in $R$.*

PROOF. This follows directly from Proposition 2.9. □

## References

Hoffman, Kenneth and Ray Kunze (1971). *Linear Algebra*. 2nd ed. Prentice-Hall. 407 pp.

Roman, Steven (2008). *Advanced Linear Algebra*. 3rd ed. Springer. 522 pp. ISBN: 978-0-387-72828-5.