LINEAR ALGEBRA

TSOGTGEREL GANTUMUR

ABSTRACT. We review some concepts from linear algebra over $\mathbb{R}$.

CONTENTS

1. LINEAR MAPPINGS AND MATRICES

The basic object here from our perspective is $\mathbb{R}^n$, whose elements are typically written as $x = (x_1, \ldots, x_n)$. We have the two basic operations $x \pm y = (x_1 \pm y_1, \ldots, x_n \pm y_n)$ and $\lambda x = (\lambda x_1, \ldots, \lambda x_n)$, where $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Any vector $x \in \mathbb{R}^n$ can be written as

$$x = x_1 e_1 + \ldots + x_n e_n, \tag{1}$$

where the so-called *standard basis vectors* $e_k \in \mathbb{R}^n$, $k = 1, 2, \ldots, n$, are defined as

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0, 0), \\ e_2 &= (0, 1, 0, \ldots, 0, 0), \\ \ldots & \quad \ldots \\ e_n &= (0, 0, 0, \ldots, 0, 1). \end{aligned} \tag{2}$$

In other words, the i-th component of the vector e_k is given by

$$(e_k)_i = \delta_{ik} \equiv \begin{cases} 1 & \text{for } i = k, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Here, the symbol δ_{ik} is called *Kronecker's delta*.

The central concept of linear algebra is that of *linear functions* (other names include linear maps, mappings, and transformations). A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called *linear* if

$$F(\alpha x + \beta y) = \alpha F(x) + \beta F(y), \tag{4}$$

for any scalars $\alpha, \beta \in \mathbb{R}$ and vectors $x, y \in \mathbb{R}^n$. In this context, $\mathbb{R}^n$ and $\mathbb{R}^m$ are called the *domain* and the *codomain* of F.

Example 1.1. The *identity map* $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\mathrm{id}(x) = x$, and the map $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F(x_1, x_2) = x_1 - 3x_2$ are linear.

Date: 2016/11/17 23:51 (EDT). Incomplete draft.

Given any $x \in \mathbb{R}^n$, by linearity, we infer

$$F(x) = F(x_1 e_1 + \ldots + x_n e_n) = x_1 F(e_1) + \ldots + x_n F(e_n). \tag{5}$$

This means that the collection of vectors $F(e_1), \ldots, F(e_n) \in \mathbb{R}^m$ completely determines the linear map F. Conversely, given a collection of vectors $f_1, \ldots, f_n \in \mathbb{R}^m$, the map $F : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$F(x) = x_1 f_1 + \ldots + x_n f_n, \tag{6}$$

is linear. Therefore, a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ can be identified with a collection of n vectors from $\mathbb{R}^m$. Furthermore, if we denote the i-the component of $F(e_k)$ or f_k by $a_{i,k}$, then

$$[F(x)]_i = a_{i,1}x_1 + a_{i,2}x_2 \ldots + a_{i,n}x_n = \sum_{k=1}^{n} a_{i,k}x_k. \tag{7}$$

The coefficients $a_{i,k}$ are usually arranged in a rectangular array, as in

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{pmatrix}, \tag{8}$$

and the whole collection A is called an $m \times n$ *matrix*. When there is no risk of confusion, we simply write a_{11}, a_{ik}, etc., instead of $a_{1,1}$ and $a_{i,k}$. Thus, any $m \times n$ matrix generates a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$, and any linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ corresponds to an $m \times n$ matrix. In other words, denoting the space of $m \times n$ matrices by $\mathbb{R}^{m\times n}$, and the set of all linear maps between $\mathbb{R}^n$ and $\mathbb{R}^m$ by $L(\mathbb{R}^n, \mathbb{R}^m)$, there is an identification $L(\mathbb{R}^n, \mathbb{R}^m) = \mathbb{R}^{m\times n}$.

Example 1.2. (a) For the identity map $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$, we have $\mathrm{id}(e_k) = e_k$, $k = 1, \ldots, n$, and hence the matrix entries are $[\mathrm{id}(e_k)]_i = (e_k)_i = \delta_{ik}$. This matrix

$$I = I_{n\times n} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{pmatrix}, \tag{9}$$

is called the $n \times n$ *identity matrix*.

(b) For the zero map $F : \mathbb{R}^n \to \mathbb{R}^m$ defined by $F(x) = 0$ for all $x \in \mathbb{R}^n$, the corresponding matrix will obviously be the $m \times n$ matrix whose entries are all 0. This matrix is called the $m \times n$ *zero matrix*, and denoted by simply 0, or by $0_{m\times n}$.

(c) For $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $F(x_1, x_2) = (x_1 + 3x_2, -x_1)$, the corresponding matrix is

$$A = \begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix}. \tag{10}$$

(d) For $F : \mathbb{R} \to \mathbb{R}^2$ given by $F(t) = (2t, 5t)$, the corresponding matrix is

$$A = \begin{pmatrix} 2 \\ 5 \end{pmatrix}. \tag{11}$$

For $A, C \in \mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$, the operations $A \pm C$ and λA are defined in the same way as for vectors, by identifying $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{mn}$. In addition, a multiplicative operation for matrices arises in connection with composition of linear maps, as follows. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^\ell \to \mathbb{R}^n$ be linear maps. Then the composition $F \circ G : \mathbb{R}^\ell \to \mathbb{R}^m$ defined by $(F \circ G)(x) = F(G(x))$ is linear, because

$$F\big(G(\alpha x + \beta y)\big) = F\big(\alpha G(x) + \beta G(y)\big) = \alpha F\big(G(x)\big) + \beta F\big(G(y)\big). \tag{12}$$

Suppose that $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ and $B = (b_{jk}) \in \mathbb{R}^{n\times k}$ are the matrices corresponding to F and G, respectively. Then we have

$$F\big(G(e_k)\big) = F(b_{1k}e_1 + \ldots + b_{nk}e_n) = b_{1k}F(e_1) + \ldots + b_{nk}F(e_n), \tag{13}$$

and hence the (i,k)-th entry of the matrix corresponding to $F\circ G$ is

$$[F\big(G(e_k)\big)]_i = b_{1k}[F(e_1)]_i + \ldots + b_{nk}[F(e_n)]_i = a_{i1}b_{1k} + \ldots + a_{in}b_{nk}. \tag{14}$$

In view of this, we define the *product* of A and B as the matrix $AB \in \mathbb{R}^{m\times\ell}$, whose entries are given by

$$[AB]_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} = a_{i1}b_{1k} + \ldots + a_{in}b_{nk}. \tag{15}$$

Exercise 1.1. Prove that matrix multiplication is associative ($(AB)C = A(BC)$) and distributive ($A(B + D) = AB + AD$), but *not* commutative in general.

Recall that the components of $F(x)$ for $x \in \mathbb{R}^n$ is given by (7). In light of the definition of matrix multiplication, (7) tells us that $F(x)$ is simply the product of the matrices A and x, if we consider x as an $n \times 1$ matrix, i.e., as a *column vector*. Thus we define the *(matrix-vector) product* between a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^n$ as the vector $Ax \in \mathbb{R}^m$, whose components are given by

$$[Ax]_i = \sum_{k=1}^{n} a_{ik}x_k = a_{i1}x_1 + \ldots + a_{in}x_n. \tag{16}$$

With this definition at hand, we can write $F(x) = Ax$ for $x \in \mathbb{R}^n$.

Remark 1.3. Given a vector $x \in \mathbb{R}^n$, we can define the linear map $G : \mathbb{R} \to \mathbb{R}^n$ by $G(t) = tx$, so that the matrix of G is an $n \times 1$ matrix with the entries identical to the components of x. Let us denote the matrix of G by $X \in \mathbb{R}^{n\times 1}$. Then the matrix of the map $F\circ G$ is $AX \in \mathbb{R}^{m\times 1}$, meaning that we have $(F \circ G)(t) = ty$ for some $y \in \mathbb{R}^m$. This vector y is precisely what we defined as the matrix-vector product Ax.

2. Linear spaces

Given a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ with the associated matrix $A \in \mathbb{R}^{m\times n}$, we define the *range of* F or the *range of* A as

$$\mathrm{ran}(F) = \mathrm{ran}(A) = \{Ax : x \in \mathbb{R}^n\} \subset \mathbb{R}^m. \tag{17}$$

This set is also called the *image* of A or of F, or the *column space* of A. Furthermore, we define the *kernel of* F or the *kernel of* A as

$$\ker(F) = \ker(A) = \{x \in \mathbb{R}^n : Ax = 0\} \subset \mathbb{R}^n, \tag{18}$$

Other names for this set include the *null space* of A, and the *zero set* of F. The range and the kernel of F contain important structural information on the map F. This can be gleaned by considering the equation $F(x) = b$, where $b \in \mathbb{R}^m$ is given. It is clear that $F(x) = b$ has a solution if and only if $b \in \mathrm{ran}(F)$. Moreover, if $F(x) = F(y) = b$, then $F(x - y) = F(x) - F(y) = 0$, and hence $x - y \in \ker(F)$. On the other hand, if $F(x) = b$ and $h \in \ker F$, then $F(x + h) = F(x) + F(h) = b$. This means that the extent of non-uniqueness of the solutions of $F(x) = b$ is precisely measured by the kernel of F. In particular, $F(x) = b$ has a unique solution if and only if $b \in \mathrm{ran}(F)$ and $\ker(F) = \{0\}$.

Let $x, y \in \ker(A)$ and $\alpha, \beta \in \mathbb{R}$. Then we have

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay = 0, \tag{19}$$

that is, $\alpha x + \beta y \in \ker(A)$. Let $\xi, \eta \in \operatorname{ran}(A)$, that is, $\xi = Ax$ and $\eta = Ay$ for some $x, y \in \mathbb{R}^n$. Then for $\alpha, \beta \in \mathbb{R}$, we have

$$\alpha\xi + \beta\eta = \alpha Ax + \beta Ay = A(\alpha x + \beta y), \tag{20}$$

that is, $\alpha\xi + \beta\eta \in \operatorname{ran}(A)$. The kernel and the range of A are examples of linear spaces.

Definition 2.1. A nonempty subset $V \subset \mathbb{R}^n$ is called a *linear space* (or a *linear subspace* of $\mathbb{R}^n$) if $\alpha x + \beta y \in V$ for all $x, y \in V$ and $\alpha, \beta \in \mathbb{R}$.

Let us denote the columns of A by $A_1, \ldots, A_n \in \mathbb{R}^m$. Then for $x \in \mathbb{R}^n$, the matrix-vector product Ax can be written as

$$Ax = x_1 A_1 + x_2 A_2 + \ldots + x_n A_n. \tag{21}$$

In general, the right hand side is called the *linear combination* of the vectors $A_1, \ldots, A_n \in \mathbb{R}^m$, with the *coefficients* given by $x_1, \ldots, x_n \in \mathbb{R}$. Thus the range of A,

$$\operatorname{ran}(A) = \{x_1 A_1 + \ldots + x_n A_n : x \in \mathbb{R}^n\}, \tag{22}$$

is simply the collection of all possible linear combinations of the columns of A, which explains the alternative name "column space." We say that the space $\operatorname{ran}(A)$ is *generated*, or *spanned* by the collection $\{A_1, \ldots, A_n\}$, or simply that $\operatorname{ran}(A)$ is equal to the *span* of $\{A_1, \ldots, A_n\}$, with the latter defined by

$$\operatorname{span}\{A_1, \ldots, A_n\} = \{x_1 A_1 + \ldots + x_n A_n : x \in \mathbb{R}^n\}. \tag{23}$$

The first step towards understanding $\operatorname{ran}(A)$ is to try to come up with an "efficient" description of it, by removing redundant vectors from the collection $\{A_1, \ldots, A_n\}$. In other words, we want to have a *minimal* subcollection $\{V_1, \ldots, V_k\} \subset \{A_1, \ldots, A_n\}$, with the property that

$$\operatorname{span}\{V_1, \ldots, V_k\} = \operatorname{span}\{A_1, \ldots, A_n\}. \tag{24}$$

The vector A_n can be removed from the collection $\{A_1, \ldots, A_n\}$ without changing its span if and only if A_n can be written as a linear combination of the remaining vectors $\{A_1, \ldots, A_{n-1}\}$, that is, if and only if there exist numbers $x_1, \ldots, x_{n-1} \in \mathbb{R}$ such that

$$A_n = x_1 A_1 + \ldots + x_{n-1} A_{n-1}. \tag{25}$$

We see that *some* vector A_i can be removed from the collection $\{A_1, \ldots, A_n\}$ without changing its span if and only if there exist numbers $x_1, \ldots, x_n \in \mathbb{R}$, not all equal to 0, such that

$$x_1 A_1 + \ldots + x_n A_n = 0. \tag{26}$$

In terms of A, the latter condition is the same as $\ker(A) \neq \{0\}$. This also means that we cannot remove any vector from the collection $\{V_1, \ldots, V_k\}$ without changing its span if and only if the only way for the equality

$$x_1 V_1 + \ldots + x_k V_k = 0, \tag{27}$$

to hold is to have $x_1 = \ldots = x_k = 0$. If we form the matrix $V \in \mathbb{R}^{m\times k}$ with the vectors $V_1, \ldots, V_k$ as columns, then the latter condition means that $\ker(V) = \{0\}$.

Definition 2.2. A set $\{V_1, \ldots, V_k\} \subset \mathbb{R}^m$ is called *linearly independent* if

$$x_1 V_1 + \ldots + x_k V_k = 0 \qquad \text{implies} \qquad x_1 = \ldots = x_k = 0. \tag{28}$$

It is called *linearly dependent* if there exists $x \in \mathbb{R}^k \setminus \{0\}$ with $x_1 V_1 + \ldots + x_k V_k = 0$.

By convention, the empty set is linearly independent.

Definition 2.3. A *basis* of a linear space X is a linearly independent set that spans X.

If $\{V_1, \ldots, V_k\}$ is a basis of X, then we have the following characteristic properties.

- Any $x \in X$ can be written as
$$x = \xi_1 V_1 + \ldots + \xi_k V_k, \tag{29}$$
for some $\xi \in \mathbb{R}^k$, because $\{V_1, \ldots, V_k\}$ spans X.
- This expansion is unique, in the sense that if
$$\xi_1 V_1 + \ldots + \xi_k V_k = \eta_1 V_1 + \ldots + \eta_k V_k, \tag{30}$$
for $\xi, \eta \in \mathbb{R}^k$, then $\xi = \eta$. The reason is that we can write the preceding equality as
$$(\xi_1 - \eta_1)V_1 + \ldots + (\xi_k - \eta_k)V_k = 0, \tag{31}$$
which implies $\xi = \eta$ by linear independence.

Example 2.4. (a) With the empty set considered as a subset of a linear space X, we have $\sum_{x\in\varnothing} x = 0$ by convention, so the empty set is a basis of the trivial vector space $\{0\}$.
(b) The set $\{e_1, e_2, \ldots, e_n\} \subset \mathbb{R}^n$ is a basis of $\mathbb{R}^n$, and called the *standard basis of* $\mathbb{R}^n$.

Example 2.5. Consider the matrix
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 2 & 5 & 7 \end{pmatrix}, \tag{32}$$
so that $A_1 = (1, 0, 2)$, $A_2 = (0, 1, 5)$, and $A_3 = (1, 1, 7)$. Since $A_1 + A_2 - A_3 = 0$, we can express any of A_1, A_2 and A_3 in terms of the other two.

(a) Let us remove A_3, and consider the resulting collection $\{A_1, A_2\}$. This is linearly independent, because the first two components of the equation $\alpha A_1 + \beta A_2 = 0$ already imply $\alpha = \beta = 0$. Thus we have
$$\operatorname{ran}(A) = \operatorname{span}\{A_1, A_2\}, \tag{33}$$
and the collection $\{A_1, A_2\}$ cannot be reduced any further, meaning that $\{A_1, A_2\}$ is a basis of $\operatorname{ran}(A)$.
(b) Let us remove A_1, and consider the resulting collection $\{A_2, A_3\}$. This is linearly independent, because the first two components of the equation $\alpha A_2 + \beta A_3 = 0$ give $\beta = 0$ and $\alpha + \beta = 0$, yielding $\alpha = \beta = 0$. Thus $\{A_2, A_3\}$ is a basis of $\operatorname{ran}(A)$.

Suppose that we started with the collection $\{A_1, \ldots, A_n\}$, and by removing redundant vectors one after another, we ended up with the subcollection $\{V_1, \ldots, V_k\}$ that is a basis of $\operatorname{ran}(A)$. We have seen in Example 2.5 that different choices in the intermediate steps can result in a different subcollection, say, $\{W_1, \ldots, W_r\} \subset \{A_1, \ldots, A_n\}$. It is also conceivable that a basis $\{W_1, \ldots, W_r\}$ of $\operatorname{ran}(A)$ that is not necessarily a subcollection of $\{A_1, \ldots, A_n\}$ can be obtained by some other means. In any case, we prove below that the number of elements in a basis is a quantity that depends only on the linear space itself, rather than on what particular vectors we have in the basis. Thus in our situation, it is necessarily true that $r = k$.

We first prove the following fundamental result.

Theorem 2.6 (Steinitz exchange lemma)**.** *Let $\{V_1, \ldots, V_k\} \subset \mathbb{R}^n$ be a linearly independent set, and let $\{W_1, \ldots, W_r\} \subset \mathbb{R}^n$ be such that $V_1, \ldots, V_k \in \operatorname{span}\{W_1, \ldots, W_r\}$. Then $k \le r$, and after a possible reordering of the set $\{W_1, \ldots, W_r\}$, we have*
$$\operatorname{span}\{W_1, \ldots, W_r\} = \operatorname{span}\{V_1, \ldots, V_k, W_{k+1}, \ldots, W_r\}. \tag{34}$$

Proof. Since $V_1 \in Y := \operatorname{span}\{W_1, \ldots, W_r\}$, we can write
$$V_1 = \alpha_1 W_1 + \ldots + \alpha_r W_r. \tag{35}$$

If $\alpha_1 = \ldots = \alpha_r = 0$, then $V_1 = 0$, which would contradict the fact that $\{V_1\}$ is linearly independent. Hence there is at least one nonzero coefficient among $\alpha_1, \ldots, \alpha_r$. Without loss of generality, let $\alpha_1 \neq 0$. Then we have

$$W_1 = \tfrac{1}{\alpha_1} V_1 - \tfrac{\alpha_2}{\alpha_1} W_2 - \ldots - \tfrac{\alpha_r}{\alpha_1} W_r, \tag{36}$$

and hence $Y = \operatorname{span}\{V_1, W_2, \ldots, W_r\}$. Thus, we can write

$$V_2 = \alpha_1 V_1 + \alpha_2 W_2 + \ldots + \alpha_r W_r, \tag{37}$$

with the coefficients $\alpha_1, \ldots, \alpha_r$ having (possibly) new values. If $\alpha_2 = \ldots = \alpha_r = 0$, then $V_2 = \alpha_1 V_1$, which would contradict the fact that $\{V_1, V_2\}$ is linearly independent. Therefore there is at least one nonzero coefficient among $\alpha_2, \ldots, \alpha_r$. Without loss of generality, let $\alpha_2 \neq 0$. Then we have

$$W_2 = -\tfrac{\alpha_1}{\alpha_2} V_1 + \tfrac{1}{\alpha_2} V_2 - \tfrac{\alpha_3}{\alpha_2} W_3 - \ldots - \tfrac{\alpha_r}{\alpha_2} W_r, \tag{38}$$

and hence $Y = \operatorname{span}\{V_1, V_2, W_3, \ldots, W_r\}$. We can repeat this process, and end up with the conclusion $Y = \operatorname{span}\{V_1, \ldots, V_k, W_{k+1}, \ldots, W_r\}$, which also shows that $k \leq r$. □

By applying the preceding theorem with $\{W_1, \ldots, W_r\}$ given by the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$, we get the following important corollary.

Corollary 2.7 (Basis completion)**.** *Let* $\{V_1, \ldots, V_k\} \subset \mathbb{R}^n$ *be a linearly independent set. Then there exist vectors* $V_{k+1}, \ldots, V_n \in \mathbb{R}^n$ *such that* $\operatorname{span}\{V_1, \ldots, V_n\} = \mathbb{R}^n$.

In the Theorem 2.6, if $\{W_1, \ldots, W_r\}$ is a basis of $\operatorname{span}\{V_1, \ldots, V_k\}$, we can switch the roles of $\{V_1, \ldots, V_k\}$ and $\{W_1, \ldots, W_r\}$ in the entire argument, and get $r \leq k$ as well.

Corollary 2.8 (Dimension theorem)**.** *Let* $\{V_1, \ldots, V_k\} \subset \mathbb{R}^n$ *and* $\{W_1, \ldots, W_r\} \subset \mathbb{R}^n$ *be linearly independent sets such that* $\operatorname{span}\{V_1, \ldots, V_k\} = \operatorname{span}\{W_1, \ldots, W_r\}$. *Then* $k = r$.

This corollary motivates the following definition.

Definition 2.9. If a linear space X has a basis with k elements, we call k the *dimension* of X, and write $k = \dim X$.

Example 2.10. (a) We have $\dim\{0\} = 0$.
(b) For A from Example 2.5, we have $\dim \operatorname{ran}(A) = 2$.
(c) Since the standard basis $\{e_1, e_2, \ldots, e_n\} \subset \mathbb{R}^n$ has n elements, we have $\dim \mathbb{R}^n = n$.
(d) If $X \subset \mathbb{R}^n$ is a linear space and $k = \dim X$, then Corollary 2.7 implies that $k \leq n$.

Intuitively, $\dim X$ is the number of degrees of freedom in X, or how many "free variables" we need in order to describe a point in X. At this point, $\dim X$ is defined only when X admits a basis. However, the following result guarantees that this does not pose any restriction.

Theorem 2.11 (Basis theorem)**.** *Every linear space* $X \subset \mathbb{R}^n$ *admits a basis.*

Proof. Since the empty set is a basis of $\{0\}$, we can assume that $X \neq \{0\}$. Hence there is a vector $V_1 \in X$ with $V_1 \neq 0$. If $X = \operatorname{span}\{V_1\}$, we are done. If $X \neq \operatorname{span}\{V_1\}$, then there is a vector $V_2 \in X$ with $V_2 \notin \operatorname{span}\{V_1\}$. Suppose that we continued this process, and got $X = \operatorname{span}\{V_1, V_2, \ldots, V_k\}$ for some k. By construction, the set $\{V_1, V_2, \ldots, V_k\}$ is linearly independent, and so this set is a basis of X. Anticipating a contradiction, now suppose that we never get $X = \operatorname{span}\{V_1, V_2, \ldots, V_k\}$ for any k, so that we have an infinite sequence of vectors $V_1, V_2, \ldots$ in X. By construction, $\{V_1, \ldots, V_{n+1}\}$ is linearly independent, and so $\dim Y = n+1$ for $Y = \operatorname{span}\{V_1, \ldots, V_{n+1}\}$. However, since $V_i \in \mathbb{R}^n$ for all i, we have $Y \subset \mathbb{R}^n$, implying that $\dim Y \leq n$, a contradiction. □

Exercise 2.1. Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^n$ be linear spaces satisfying $\dim X = \dim Y$ and $X \subset Y$. Show that $X = Y$.

3. The rank-nullity theorem

Our quest for understanding the range and the kernel of a matrix $A \in \mathbb{R}^{m\times n}$ has lead us to the concepts of linear spaces, linear independence, span, basis, and eventually to the dimension theorem. With these concepts at hand, we now want to get back to the study of $\operatorname{ran}(A)$ and $\ker(A)$ proper. In the preceding section, we have encountered the following argument. Denoting the columns of A by $A_1, \ldots, A_n \in \mathbb{R}^m$, the columns of A are linearly independent if and only if the equality

$$Ax = x_1A_1 + x_2A_2 + \ldots + x_nA_n = 0, \tag{39}$$

implies $x_1 = \ldots = x_n = 0$. In other words, $\dim \operatorname{ran}(A) = n$ if and only if $\ker(A) = \{0\}$. This suggests that there might be some law relating the dimensions of $\operatorname{ran}(A)$ and $\ker(A)$.

Suppose that the columns $A_1, \ldots, A_n$ are linearly dependent, and that we remove redundant columns one by one to get a basis of $\operatorname{ran}(A)$. Let us consider the first step of this process. Thus without loss of generality, we assume that

$$\alpha_1A_1 + \alpha_2A_2 + \ldots + \alpha_nA_n = 0, \tag{40}$$

with $\alpha_1 = 1$. Taking this into account, we infer that $x \in \ker(A)$ if and only if

$$x_1A_1 + x_2A_2 + \ldots + x_nA_n = (x_2 - \alpha_2x_1)A_2 + \ldots + (x_n - \alpha_nx_1)A_n = 0. \tag{41}$$

Let us denote by A' the matrix consisting of the columns $A_2, \ldots, A_n$, that is, the matrix that results from removing the column A_1 from A. Recall that $\operatorname{ran}(A) = \operatorname{ran}(A')$ by construction. By the aforementioned argument, we also have

$$x \in \ker(A) \qquad \iff \qquad \begin{pmatrix} x_2 - \alpha_2x_1 \\ \ldots \\ x_n - \alpha_nx_1 \end{pmatrix} \in \ker(A'). \tag{42}$$

Hence in order to specify an arbitrary element $x \in \ker(A)$, we can choose $x_1 \in \mathbb{R}$ freely, and then choose the remaining components $x_2, \ldots, x_n$ so that the second inclusion in (42) holds. This suggests that $\dim\ker(A) = \dim\ker(A') + 1$. To make this argument absolutely convincing, let $W_1, \ldots, W_r \in \mathbb{R}^{n-1}$ be a basis of $\ker(A')$, and define $V_0, V_1, \ldots, V_r \in \mathbb{R}^n$ by

$$V_0 = \begin{pmatrix} 1 \\ \alpha_2 \\ \ldots \\ \alpha_n \end{pmatrix}, \qquad V_1 = \begin{pmatrix} 0 \\ [W_1]_1 \\ \ldots \\ [W_1]_{n-1} \end{pmatrix}, \qquad \ldots, \qquad V_r = \begin{pmatrix} 0 \\ [W_r]_1 \\ \ldots \\ [W_r]_{n-1} \end{pmatrix}, \tag{43}$$

where $[W_i]_j$ denotes the j-th component of the vector $W_i \in \mathbb{R}^{n-1}$. Basically, what we have is $V_0 = \alpha$ and $V_i = (0, W_i)$ for $i = 1, \ldots, r$. The set $\{V_0, V_1, \ldots, V_r\}$ is linearly independent, and $AV_i = 0$ for $i = 0, \ldots, r$. The latter implies that $\operatorname{span}\{V_0, V_1, \ldots, V_r\} \subset \ker(A)$. Moreover, if $x \in \ker(A)$, then by (42), for some coefficients $\beta_1, \ldots, \beta_r \in \mathbb{R}$, we have

$$\begin{pmatrix} x_2 - \alpha_2x_1 \\ \ldots \\ x_n - \alpha_nx_1 \end{pmatrix} = \beta_1W_1 + \ldots + \beta_rW_r, \tag{44}$$

or, equivalently,

$$\begin{pmatrix} x_2 \\ \ldots \\ x_n \end{pmatrix} = \beta_0 \begin{pmatrix} \alpha_2 \\ \ldots \\ \alpha_n \end{pmatrix} + \beta_1W_1 + \ldots + \beta_rW_r \qquad \text{with} \qquad \beta_0 = x_1. \tag{45}$$

This precisely means that $x \in \operatorname{span}\{V_0, V_1, \ldots, V_r\}$, and hence $\{V_0, V_1, \ldots, V_r\}$ is a basis of $\ker(A)$. We conclude that

$$\dim\ker(A) = r + 1 = \dim\ker(A') + 1. \tag{46}$$

Suppose now that we started with $A^{(0)} = A$, and removed redundant columns one by one to get a sequence of matrices $A^{(0)}, A^{(1)}, \ldots, A^{(k)}$, with $A^{(k)} \in \mathbb{R}^{m\times(n-k)}$ having linearly independent columns. Since $\operatorname{ran}(A^{(k)}) = \operatorname{ran}(A)$, we have $\dim\operatorname{ran}(A) = n-k$. On the other hand, as $A^{(k)}$ has linearly independent columns, its kernel is trivial, and so (46) yields

$$\dim\ker(A) = \dim\ker(A^{(1)}) + 1 = \ldots = \dim\ker(A^{(k)}) + k = k. \tag{47}$$

This, at last, gives the sought relationship between $\dim\operatorname{ran}(A)$ and $\dim\ker(A)$, which turns out to be the following beautiful equality:

$$\dim\operatorname{ran}(A) + \dim\ker(A) = n. \tag{48}$$

We did not spend much work to prove it, but it is one of most important results from linear algebra, called the *rank-nullity theorem*. The name is explained by the following terminology.

Definition 3.1. The *rank* and the *nullity* of a matrix $A \in \mathbb{R}^{m\times n}$ are the dimension of its range and of its kernel, respectively, and they are denoted by

$$\operatorname{rank}(A) = \dim\operatorname{ran}(A), \qquad \operatorname{nullity}(A) = \dim\ker(A). \tag{49}$$

Let us state (48) in terms of rank and nullity.

Theorem 3.2 (Rank-nullity theorem)**.** *For $A \in \mathbb{R}^{m\times n}$, we have*

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n. \tag{50}$$

This theorem can be very powerful, because it reduces the complexity of studying two quantities into that of studying either of those quantities. We will see many important applications in what follows.

4. Invertible matrices

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping, with the associated matrix $A \in \mathbb{R}^{m\times n}$. If for every $\xi \in \mathbb{R}^m$ there exists a unique $x \in \mathbb{R}^n$ satisfying $F(x) = \xi$, then we say that F is *invertible*, and call the map $F^{-1} : \xi \mapsto x$ the *inverse* of F. As we have discussed in the preceding section, F is invertible if and only if $\operatorname{ran}(F) = \mathbb{R}^m$ and $\ker(F) = \{0\}$. By the rank-nullity theorem, this implies that $m + 0 = n$, or $m = n$. In other words, for a linear map to be invertible, the domain and the codomain spaces must have the same dimensions. Once $m = n$ has been established, we see that the invertibility of $F : \mathbb{R}^n \to \mathbb{R}^n$ is equivalent to either (and hence both) of the conditions $\operatorname{rank}(A) = n$ and $\operatorname{nullity}(A) = 0$.

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be invertible. Then by definition, we have $F(F^{-1}(\xi)) = \xi$ for all $\xi \in \mathbb{R}^m$. Furthermore, for each $x \in \mathbb{R}^n$, we have $F^{-1}(F(x)) = x$, because $F(y) = F(x)$ implies $y = x$. Hence for $\xi, \eta \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, we have

$$\alpha\xi + \beta\eta = \alpha F(x) + \beta F(y) = F(\alpha x + \beta y), \tag{51}$$

where $x = F^{-1}(\xi)$ and $y = F^{-1}(\eta)$, implying that

$$F^{-1}(\alpha\xi + \beta\eta) = \alpha F^{-1}(\xi) + \beta F^{-1}(\eta). \tag{52}$$

This means that the inverse map $F^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is also linear. The matrix associated to F^{-1} is called the *inverse* of A, and denoted by $A^{-1} \in \mathbb{R}^{n\times n}$. Since we have $F \circ F^{-1} = \mathrm{id}$ and $F^{-1} \circ F = \mathrm{id}$, the inverse matrix satisfies

$$AA^{-1} = I \qquad \text{and} \qquad A^{-1}A = I. \tag{53}$$

Remark 4.1. Suppose that $A, B \in \mathbb{R}^{n\times n}$ satisfy $AB = I$. This means that $ABy = y$ for all $y \in \mathbb{R}^n$, and hence $\operatorname{ran}(A) = \mathbb{R}^n$ or $\operatorname{rank}(A) = n$. Therefore, the map $F : \mathbb{R}^n \to \mathbb{R}^n$ associated to A is invertible. Moreover, (53) implies that the matrix of the inverse map is given by $A^{-1} = B$, and that $BA = I$. We can apply the same argument to the equality $BA = I$, and infer that $B^{-1} = A$.

This justifies the following definition.

Definition 4.2. A matrix $A \in \mathbb{R}^{n\times n}$ is called *invertible*, if there exists $B \in \mathbb{R}^{n\times n}$ satisfying either (and hence both) of the conditions $AB = I$ and $BA = I$. In this context, the matrix B is called the *inverse of* A, and denoted by $A^{-1} = B$.

We have previously defined A^{-1} as the matrix associated to the inverse map of $F(x) = Ax$. In view of Remark 4.1, this is equivalent to how we defined A^{-1} in Definition 4.2. Moreover, Remark 4.1 also implies that if a matrix is invertible, its inverse is unique.

Theorem 4.3. *Let $A \in \mathbb{R}^{n\times n}$. Then the following are equivalent.*

- *A is invertible.*
- $\operatorname{ran}(A) = \mathbb{R}^n$, *or* $\operatorname{rank}(A) = n$.
- $\ker(A) = \{0\}$, *or* $\operatorname{nullity}(A) = 0$.

Proof. We have actually proved this theorem in the earlier paragraphs of the current section, but we include a proof here for convenience. Suppose that A is invertible, i.e., there is $B \in \mathbb{R}^{n\times n}$ such that $AB = I$. This implies that $\operatorname{ran}(A) = \mathbb{R}^n$. If $\operatorname{ran}(A) = \mathbb{R}^n$, then by the rank-nullity theorem, we have $\ker(A) = \{0\}$.

Finally, suppose that $\ker(A) = \{0\}$. We need to show that A is invertible. First of all, the rank-nullity theorem gives $\operatorname{ran}(A) = \mathbb{R}^n$, meaning that for any $\xi \in \mathbb{R}^n$, there is $x \in \mathbb{R}^n$ such that $Ax = \xi$. In order to show that x is unique, let $Ay = Ax$, that is, $A(x - y) = 0$. Since $\ker(A) = \{0\}$, we get $x - y = 0$ or $y - x$. Thus, for any $\xi \in \mathbb{R}^n$, there exists a unique $x \in \mathbb{R}^n$ such that $Ax = \xi$. Therefore, the map $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by $F(x) = Ax$ is invertible. $\square$

Example 4.4. Let $A = (a_{ik}) \in \mathbb{R}^{n\times n}$ be a matrix satisfying the condition

$$\sum_{k\neq i} |a_{ik}| < |a_{ii}|, \qquad i = 1, \ldots, n. \tag{54}$$

We want to show that $\ker(A) = \{0\}$, and therefore A is invertible. Let $x \in \mathbb{R}^n$ with $x \neq 0$, and let $y = Ax$. Pick an index i such that $|x_i| \geq |x_k|$ for all $k = 1, \ldots, n$. Then we have

$$y_i = \sum_{k=1}^{n} a_{ik}x_k, \qquad \text{or} \qquad a_{ii}x_i = y_i - \sum_{k\neq i}^{n} a_{ik}x_k. \tag{55}$$

This implies that

$$|a_{ii}||x_i| \leq |y_i| + \sum_{k\neq i}^{n} |a_{ik}||x_k| \leq |y_i| + |x_i| \sum_{k\neq i}^{n} |a_{ik}| \tag{56}$$

and so

$$|y_i| \geq |x_i|\Big(|a_{ii}| - \sum_{k\neq i}^{n} |a_{ik}|\Big) > 0. \tag{57}$$

We conclude that $Ax \neq 0$ whenever $x \neq 0$, meaning that $\ker(A) = \{0\}$. To take a concrete example, we can immediately see that

$$A = \begin{pmatrix} 4 & 1 & 1 & 1 \\ 0 & 3 & 1 & -1 \\ 3 & -1 & -5 & 0 \\ 2 & -2 & 1 & 6 \end{pmatrix}, \tag{58}$$

is invertible.

5. Duality

Our study of matrices gained a lot from writing the matrix-vector product Ax as

$$Ax = x_1 A_1 + x_2 A_2 + \ldots + x_n A_n, \tag{59}$$

where $A_1, \ldots, A_n \in \mathbb{R}^m$ are the columns of the matrix $A \in \mathbb{R}^{m\times n}$, and $x \in \mathbb{R}^n$. Then a natural question is: What if we think of a matrix as a collection of its rows, rather than a collection of columns? This question will lead us to uncover a whole new structure known as *duality*. Let $A \in \mathbb{R}^{m\times n}$, and let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}^{1\times n}$ be the rows of A. Then we can write

$$Ax = \begin{pmatrix} \alpha_1 x \\ \ldots \\ \alpha_m x \end{pmatrix}. \tag{60}$$

The elements of $\mathbb{R}^{1\times n}$ are called *row vectors.*

Remark 5.1. It is convenient to introduce the notation $\mathbb{R}^{n*} = \mathbb{R}^{1\times n}$, and think of the elements of $\mathbb{R}^{n*}$ as n-tuples $\xi = (\xi_1, \ldots, \xi_n)$ that behave like $1 \times n$ matrices with respect to matrix-matrix and matrix-vector multiplication operations. That is, we have

$$[\xi B]_k = \xi_1 b_{1k} + \ldots + \xi_n b_{nk}, \qquad \text{and} \qquad \xi x = \xi_1 x_1 + \ldots + \xi_n x_n, \tag{61}$$

for $\xi \in \mathbb{R}^{n*}$, $B = (b_{ik}) \in \mathbb{R}^{n\times m}$, and $x \in \mathbb{R}^n$. A more abstract point of view is that the space $\mathbb{R}^{n*}$ is the set of all linear maps $f : \mathbb{R}^n \to \mathbb{R}$. Thus (60) is simply the decomposition of a linear map $F : \mathbb{R}^n \to \mathbb{R}^m$ into its components, as in $F(x) = (f_1(x), \ldots, f_m(x))$.

Remark 5.2. Given a matrix

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & \ldots & b_{1,n} \\ b_{2,1} & b_{2,2} & \ldots & b_{2,n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{m,1} & b_{m,2} & \ldots & b_{m,n} \end{pmatrix} \in \mathbb{R}^{m\times n}, \tag{62}$$

we define its *transpose* as

$$B^{\mathsf{T}} = \begin{pmatrix} b_{1,1} & b_{2,1} & \ldots & b_{m,1} \\ b_{1,2} & b_{2,2} & \ldots & b_{m,2} \\ \ldots & \ldots & \ldots & \ldots \\ b_{1,n} & b_{2,n} & \ldots & b_{m,n} \end{pmatrix} \in \mathbb{R}^{n\times m}. \tag{63}$$

For example, we have

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 3 \\ 2 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 3 & 2 \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} 5 & 7 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}. \tag{64}$$

In particular, under transposition, row vectors become column vectors and vice versa.

Exercise 5.1. Show that $(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$ for $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$.

The first thing to notice from (60) is that $x \in \ker(A)$ if and only if $\alpha_i x = 0$ for all $i = 1, \ldots, m$. In other words, $x \in \ker(A)$ if and only if $\xi x = 0$ for all $\xi \in \operatorname{coran}(A)$, where

$$\operatorname{coran}(A) = \{\eta_1 \alpha_1 + \ldots + \eta_m \alpha_m : \eta_1, \ldots, \eta_m \in \mathbb{R}\} \subset \mathbb{R}^{n*}, \tag{65}$$

is called the *corange of* A (also called the *coimage of* A, or the *row space of* A). Since the rows of A are the columns of A^{T}, the corange of A is basically the range of A:

$$\xi \in \operatorname{coran}(A) \qquad \Longleftrightarrow \qquad \xi^{\mathsf{T}} \in \operatorname{ran}(A^{\mathsf{T}}). \tag{66}$$

In particular, the dimension of $\operatorname{coran}(A)$, or the *row-rank of* A, is equal to the rank of A^{T}. Furthermore, we define the *cokernel of* A, or the *left-null space of* A, by

$$\operatorname{coker}(A) = \{\eta \in \mathbb{R}^{m*} : \eta A = 0\} \subset \mathbb{R}^{m*}. \tag{67}$$

As $(\eta A)^{\mathsf{T}} = A^{\mathsf{T}}\eta^{\mathsf{T}}$, we have the following characterization

$$\eta \in \operatorname{coker}(A) \qquad \Longleftrightarrow \qquad \eta^{\mathsf{T}} \in \ker(A^{\mathsf{T}}), \tag{68}$$

that is, the cokernel of A is the kernel of A^{T} under transposition. In particular, the dimension of $\operatorname{coker}(A)$, or the *left-nullity of* A, is equal to the nullity of A^{T}. Then the rank-nullity theorem applied to A^{T} yields

$$\dim \operatorname{coran}(A) + \dim \operatorname{coker}(A) = m. \tag{69}$$

With $A_1, \ldots, A_n$ being the columns of A, we can write

$$\eta A = (\eta A_1, \ldots, \eta A_n), \tag{70}$$

implying that $\eta \in \operatorname{coker}(A)$ if and only if $\eta x = 0$ for all $x \in \operatorname{ran}(A)$.

Since $\mathbb{R}^{n*}$ is really a space of n-tuples, it makes sense to talk about linear subspaces of $\mathbb{R}^{n*}$. Namely, a nonempty subset $V \subset \mathbb{R}^{n*}$ is called a *linear space* (or a *linear subspace* of $\mathbb{R}^{n*}$) if $\lambda\xi + \mu\eta \in V$ for all $\xi, \eta \in V$ and $\lambda, \mu \in \mathbb{R}$. In light of this definition, we see that $\operatorname{coran}(A)$ is a linear subspace of $\mathbb{R}^{n*}$, and $\operatorname{coker}(A)$ is a linear subspace of $\mathbb{R}^{m*}$.

Definition 5.3. Given a linear subspace $V \in \mathbb{R}^n$, the *annihilator of* V is

$$V^\circ = \{\xi \in \mathbb{R}^{n*} : \xi x = 0 \text{ for all } x \in V\}. \tag{71}$$

Similarly, for a linear subspace $W \in \mathbb{R}^{n*}$, the *annihilator of* W is

$$W^\circ = \{x \in \mathbb{R}^{n} : \xi x = 0 \text{ for all } \xi \in W\}. \tag{72}$$

Both V° and W° are linear spaces, and we have $V^{\circ\circ} = V$ and $W^{\circ\circ} = W$. The following result was derived earlier in this section, which we restate in terms of annihilators.

Theorem 5.4 (Corange and cokernel)**.** *For any matrix $A \in \mathbb{R}^{m\times n}$, we have*

$$\operatorname{coran}(A) = \ker(A)^\circ, \qquad \textit{and} \qquad \operatorname{coker}(A) = \operatorname{ran}(A)^\circ. \tag{73}$$

The following theorem is of fundamental importance.

Theorem 5.5 (Row rank equals column rank)**.** *For any matrix $A \in \mathbb{R}^{m\times n}$, we have*

$$\dim \operatorname{coran}(A) = \dim \operatorname{ran}(A). \tag{74}$$

Proof. Let $\operatorname{rank}(A) = k$, and let $B_1, \ldots, B_k \in \mathbb{R}^m$ be a basis of $\operatorname{ran}(A)$. Then any column A_i of A can be written as

$$A_i = c_{i1}B_1 + \ldots + c_{ik}B_k = BC_i, \tag{75}$$

where $C_i = (c_{i1}, \ldots, c_{ik}) \in \mathbb{R}^k$, and $B \in \mathbb{R}^{m\times k}$ is the matrix with $B_1, \ldots, B_k$ as its columns. In other words, we have

$$A = BC, \tag{76}$$

where $C \in \mathbb{R}^{k\times n}$ is the matrix with $C_1, \ldots, C_n$ as its columns. Denoting by $\alpha_1, \ldots, \alpha_m \in \mathbb{R}^{n*}$ the rows of A, then we infer

$$\alpha_j = b_{j1}\gamma_1 + \ldots + b_{jk}\gamma_k, \tag{77}$$

where $b_{j\ell}$ are the entries of B, and $\gamma_1, \ldots, \gamma_k \in \mathbb{R}^{n*}$ are the rows of C. This implies that $\operatorname{coran}(A) \subset \operatorname{span}\{\gamma_1, \ldots, \gamma_k\}$, and hence $\dim \operatorname{coran}(A) \leq k$ by the Steinitz exchange lemma.

Now we apply the preceding argument to A^{T}, and infer that $\dim \operatorname{coran}(A^{\mathsf{T}}) \leq \dim \operatorname{ran}(A^{\mathsf{T}})$. Since $\dim \operatorname{coran}(A^{\mathsf{T}}) = \dim \operatorname{ran}(A^{\mathsf{TT}}) = \dim \operatorname{ran}(A) = k$ and $\dim \operatorname{ran}(A^{\mathsf{T}}) = \dim \operatorname{coran}(A)$, we get $k \leq \dim \operatorname{coran}(A)$, completing the proof. $\square$

6. Determinants

As linear independence is such an important concept, it would be convenient if we were able to detect whether or not a set of vectors is linearly independent simply by computing a function of the vectors. That is, it would be great if we had a function ω of k arguments $V_1, \ldots, V_k \in \mathbb{R}^n$, with the property that

$$\omega(V_1, \ldots, V_k) = 0 \qquad \Longleftrightarrow \qquad \{V_1, \ldots, V_k\} \text{ is linearly dependent.} \tag{78}$$

From the outset, we restrict ourselves to the linear setting. Thus, we look for a function

$$\omega : \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}, \tag{79}$$

that is linear in each of its arguments, meaning that

$$\omega(\ldots, V_{i-1}, \lambda V + \mu W, V_{i+1}, \ldots) = \lambda\omega(\ldots, V_{i-1}, V, V_{i+1}, \ldots) + \mu\omega(\ldots, V_{i-1}, W, V_{i+1}, \ldots), \tag{80}$$

for all $V_1, \ldots, V_k, V, W \in \mathbb{R}^n$, and $\lambda, \mu \in \mathbb{R}$. Such functions are called *multilinear* or *k-linear forms*. Since we want $\omega(V_1, \ldots, V_k)$ to be zero if $\{V_1, \ldots, V_k\}$ is linearly dependent, a minimal requirement would be to have it zero when $V_i = V_j$ for some $i \neq j$, i.e.,

$$\omega(\ldots, V, \ldots, V, \ldots) = 0 \qquad \text{for any} \quad V \in \mathbb{R}^n. \tag{81}$$

Multilinear forms satisfying (81) are called *alternating k-linear forms.*

Example 6.1. The requirement (81) is vacuous when $k = 1$, and therefore any linear function $\omega : \mathbb{R}^n \to \mathbb{R}$ is an alternating 1-linear form (or simply linear form). An example of an alternating 2-linear form (or bilinear form) in $\mathbb{R}^3$ is given by $\omega(V, W) = v_1w_3 - v_3w_1$ for $V = (v_1, v_2, v_3) \in \mathbb{R}^3$ and $W = (w_1, w_2, w_3) \in \mathbb{R}^3$. With this form, note that even though the vectors $\xi = (1, 0, 0)$ and $\eta = (0, 1, 0)$ are linearly independent, we have $\omega(\xi, \eta) = 0$. However, $\tilde{\omega}(V, W) = v_1w_2 - v_2w_1$ is an alternating bilinear form in $\mathbb{R}^3$ with $\tilde{\omega}(\xi, \eta) = 1$.

Remark 6.2. The preceding example shows that there might not be a single alternating k-linear form ω that can detect linearly independent vectors as in (78). However, it turns out that the collection of *all* alternating k-linear forms is adequate for the task. Suppose that $\{V_1, \ldots, V_k\}$ is a *linearly dependent* set, and let ω be an alternating k-linear form. Without loss of generality, assume that $V_1 = \alpha_2V_2 + \ldots + \alpha_kV_k$. Then we have

$$\omega(V_1, V_2, \ldots, V_k) = \alpha_2\omega(V_2, V_2, \ldots, V_k) + \ldots + \alpha_k\omega(V_k, V_2, \ldots, V_k) = 0, \tag{82}$$

implying that if $\{V_1, \ldots, V_k\}$ linearly dependent, then $\omega(V_1, V_2, \ldots, V_k) = 0$ for any alternating k-linear form ω. In the other direction, we will show later in Theorem 6.5 that if $\{V_1, \ldots, V_k\}$ is *linearly independent*, then there is at least one alternating k-linear form ω such that $\omega(V_1, V_2, \ldots, V_k) \neq 0$.

With the notation $\alpha(V, W) = \omega(\ldots, V + W, \ldots, V + W, \ldots)$, we have

$$\begin{aligned} \alpha(V + W, V + W) &= \alpha(V, V) + \alpha(V, W) + \alpha(W, V) + \alpha(W, W) \\ &= \alpha(V, V) + \alpha(W, W), \end{aligned} \tag{83}$$

implying that ω is *antisymmetric* in each pair of its arguments:

$$\omega(\ldots, V, \ldots, W, \ldots) + \omega(\ldots, W, \ldots, V, \ldots) = 0 \qquad \text{for any} \quad V, W \in \mathbb{R}^n. \tag{84}$$

Exercise 6.1. Show that if ω as in (79) is linear in its *first* argument, and if it satisfies the antisymmetry condition (84), then ω is an alternating k-linear form.

Let ω be an alternating k-linear form, and define the coefficients

$$\omega_{i_1\dots i_k} = \omega(e_{i_1}, \dots, e_{i_k}), \qquad i_1, \dots, i_k = 1, \dots, n, \tag{85}$$

where $e_j \in \mathbb{R}^n$ are the standard basis vectors. If $V_i = v_{i1}e_1 + \dots + v_{in}e_n$ for $i = 1, \dots, k$, then

$$\omega(V_1, \dots, V_k) = \omega\Big(\sum_{i_1=1}^{n} v_{1i_1}e_{i_1}, \dots, \sum_{i_k=1}^{n} v_{ki_k}e_{i_k}\Big) = \sum_{i_1,\dots,i_k=1}^{n} \omega_{i_1\dots i_k} v_{1i_1}\cdots v_{ki_k}, \tag{86}$$

meaning that the coefficients $\{\omega_{i_1\dots i_k}\}$ determine the multilinear form ω completely. Furthermore, the antisymmetry condition (84) implies

$$\omega_{\dots i\dots j\dots} = \omega(\dots, e_i, \dots, e_j, \dots), = -\omega(\dots, e_j, \dots, e_i, \dots) = -\omega_{\dots j\dots i\dots}, \tag{87}$$

that is, $\omega_{i_1\dots i_k}$ is antisymmetric in each pair of its indices. In particular, we have $\omega_{i_1\dots i_k} = 0$ whenever $i_a = i_b$ for $a \neq b$.

Now suppose that $\{\omega_{i_1\dots i_k} : i_1, \dots, i_k = 1, \dots, n\}$ is a collection of coefficients satisfying

$$\omega_{\dots i\dots j\dots} = -\omega_{\dots j\dots i\dots}. \tag{88}$$

Then $\omega : \mathbb{R}^n \times \dots \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\omega(V_1, \dots, V_k) = \sum_{i_1,\dots,i_k=1}^{n} \omega_{i_1\dots i_k} v_{1i_1}\cdots v_{ki_k}, \tag{89}$$

is an alternating k-linear form. Thus we have established a one-to-one correspondence between alternating k-linear forms and the coefficients $\{\omega_{i_1\dots i_k} : i_1, \dots, i_k = 1, \dots, n\}$ satisfying (88). At this juncture, a crucial question is if and "how many" nontrivial alternating k-linear forms exist. The set of all possible coefficients $\{\omega_{i_1\dots i_k} : i_1, \dots, i_k = 1, \dots, n\}$ is the same as $\mathbb{R}^{nk}$, but the conditions (88) significantly reduce the possibilities. Namely, by permuting indices and using (88), all coefficients can be expressed only in terms of the coefficients $\omega_{i_1\dots i_k}$ with $i_1 < i_2 < \dots < i_k$. For example, all alternating 2-linear forms in $\mathbb{R}^3$ can be generated by specifying the three coefficients ω_{12}, ω_{13}, and ω_{23}, which yields

$$\omega(V, W) = \sum_{i,j=1}^{n} \omega_{ij} v_i w_j = \omega_{12}(v_1w_2 - v_2w_1) + \omega_{13}(v_1w_3 - v_3w_1) + \omega_{23}(v_2w_3 - v_3w_2). \tag{90}$$

In general, the dimension of $\{\omega_{i_1\dots i_k} \in \mathbb{R} : 1 \leq i_1 < i_2 < \dots < i_k \leq n\}$ is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

A particular case of interest is alternating n-linear forms in $\mathbb{R}^n$, where we have only one independent component in the coefficients $\omega_{i_1\dots i_n}$, because there is only one possibility to have n integers satisfying $1 \leq i_1 < i_2 < \dots < i_n \leq n$. Thus, if $(i_1, \dots, i_n)$ is an even permutation of $(1, \dots, n)$, then $\omega_{i_1\dots i_n} = \omega_{1\dots n}$, and if $(i_1, \dots, i_n)$ is an odd permutation of $(1, \dots, n)$, then $\omega_{i_1\dots i_n} = -\omega_{1\dots n}$. In other words, taking into account that $\omega_{1\dots n} = \omega(e_1, \dots, e_n)$, any alternating n-linear form can be written as

$$\omega(V_1, \dots, V_n) = \omega(e_1, \dots, e_n) \sum_{i_1,\dots,i_n=1}^{n} \operatorname{sign}(i_1, \dots, i_n)\, v_{1i_1}\cdots v_{ni_n}, \tag{91}$$

where

$$\operatorname{sign}(i_1, \dots, i_n) = \begin{cases} 1 & \text{if } (i_1, \dots, i_n) \text{ is an even permutation of } (1, \dots, n), \\ -1 & \text{if } (i_1, \dots, i_n) \text{ is an odd permutation of } (1, \dots, n), \\ 0 & \text{if } (i_1, \dots, i_n) \text{ is not a permutation of } (1, \dots, n). \end{cases} \tag{92}$$

For example, we have $\operatorname{sign}(1,2) = 1$, $\operatorname{sign}(2,1) = -1$, and $\operatorname{sign}(1,1) = \operatorname{sign}(2,2) = 0$. In light of (91), the condition $\omega(e_1, \dots, e_n) = 1$ uniquely defines an alternating n-linear form, the form called the n-dimensional *determinant*:

$$\det(V_1, \ldots, V_n) = \sum_{i_1, \ldots, i_n = 1}^{n} \operatorname{sign}(i_1, \ldots, i_n)\, v_{1 i_1} \cdots v_{n i_n}. \tag{93}$$

The determinant is customarily thought of as an n-linear form applied to the columns of an $n \times n$ matrix, and written as

$$\det(B) = \sum_{i_1, \ldots, i_n = 1}^{n} \operatorname{sign}(i_1, \ldots, i_n)\, b_{i_1 1} \cdots b_{i_n n}, \qquad B = (b_{ik}) \in \mathbb{R}^{n \times n}. \tag{94}$$

Example 6.3. For 1×1 matrices, we have $\det(b) = b$. The 2-dimensional determinant is

$$\begin{aligned} \det \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} &= \operatorname{sign}(1,1) b_{11} b_{12} + \operatorname{sign}(1,2) b_{11} b_{22} + \operatorname{sign}(2,1) b_{21} b_{12} + \operatorname{sign}(2,2) b_{21} b_{22} \\ &= b_{11} b_{22} - b_{21} b_{22} = b_{11} \det(b_{22}) - b_{21} \det(b_{22}). \end{aligned} \tag{95}$$

Furthermore, the 3-dimensional case is

$$\begin{aligned} \det \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} &= b_{11} b_{22} b_{33} - b_{11} b_{32} b_{23} - b_{21} b_{12} b_{33} \\ &\quad + b_{21} b_{32} b_{13} + b_{31} b_{12} b_{23} - b_{31} b_{22} b_{13} \\ &= b_{11}(b_{22} b_{33} - b_{32} b_{23}) - b_{21}(b_{12} b_{33} - b_{32} b_{13}) + b_{31}(b_{12} b_{23} - b_{22} b_{13}) \\ &= b_{11} \det \begin{pmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{pmatrix} - b_{21} \det \begin{pmatrix} b_{12} & b_{13} \\ b_{32} & b_{33} \end{pmatrix} + b_{31} \det \begin{pmatrix} b_{12} & b_{13} \\ b_{22} & b_{23} \end{pmatrix}. \end{aligned} \tag{96}$$

Example 6.4. Let $B \subset \mathbb{R}^{n \times n}$ be an *upper triangular* matrix, in the sense that $b_{ik} = 0$ whenever $i > k$. This means that the indices of the nonzero terms in the sum (94) must satisfy $i_k \leq k$. Since the only permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ with $i_k \leq k$ is $(1, \ldots, n)$, the determinant of a triangular matrix equals the product of the entries on the main diagonal:

$$\det(B) = b_{11} b_{22} \cdots b_{nn}. \tag{97}$$

Exercise 6.2. Show that $\det(A) = \det(A^{\mathsf{T}})$.

The property (91) can now be rephrased as

$$\omega(V_1, \ldots, V_n) = \omega(e_1, \ldots, e_n) \det(V_1, \ldots, V_n), \tag{98}$$

for any alternating n-linear form ω. An important application of this is as follows. Given any matrix $A \in \mathbb{R}^{n \times n}$, the formula

$$\omega(V_1, \ldots, V_n) = \det(AV_1, \ldots, BV_n), \tag{99}$$

defines an alternating n-linear form, with $\omega(e_1, \ldots, e_n) = \det(A)$. Hence by (98), we infer

$$\det(AV_1, \ldots, AV_n) = \det(A) \det(V_1, \ldots, V_n), \tag{100}$$

or simply

$$\det(AB) = \det(A) \det(B), \tag{101}$$

for any $A, B \in \mathbb{R}^{n \times n}$.

Exercise 6.3. Show that if A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.

Theorem 6.5. *A set $\{V_1, \ldots, V_k\}$ of vectors in $\mathbb{R}^n$ is linearly dependent if and only if $\omega(V_1, \ldots, V_k) = 0$ for any alternating k-linear form ω.*

Proof. We have shown in Remark 6.2 that if $\{V_1, \ldots, V_k\}$ is a linearly dependent set and if ω is any alternating k-linear form, then $\omega(V_1, \ldots, V_k) = 0$. In the other direction, let $\{V_1, \ldots, V_k\}$ be a *linearly independent* set. We want to show that there is at least one alternating k-linear form ω such that $\omega(V_1, \ldots, V_k) \neq 0$. We invoke Corollary 2.7 into complete the set $\{V_1, \ldots, V_k\}$ to a basis $\{V_1, \ldots, V_n\}$ of $\mathbb{R}^n$. This basis defines the invertible linear map $F : \mathbb{R}^n \to \mathbb{R}^n$ as

$$F(x) = x_1 V_1 + \ldots + x_n V_n, \tag{102}$$

and the inverse F^{-1} is the map that sends

$$W = x_1 V_1 + \ldots + x_n V_n, \tag{103}$$

to the coefficient vector $x \in \mathbb{R}^n$. Let $A \in \mathbb{R}^{n\times n}$ be the matrix associated to F^{-1}, and let $B \in \mathbb{R}^{k\times n}$ be the matrix formed by the first k rows of A. In other words, $BW \in \mathbb{R}^k$ is equal to the vector $(x_1, \ldots, x_k)$ in column format, for $W \in \mathbb{R}^n$. Finally, we let

$$\omega(W_1, \ldots, W_k) = \det(BW_1, \ldots, BW_k), \tag{104}$$

which defines an alternating k-linear form in $\mathbb{R}^n$. Then we have

$$\omega(V_1, \ldots, V_k) = \det(BV_1, \ldots, BV_k) = \det(e_1, \ldots, e_k) = 1, \tag{105}$$

where $\{e_1, \ldots, e_k\}$ is the standard basis of $\mathbb{R}^k$. ☐

Since up to scaling, the determinant is the only alternating n-linear form in $\mathbb{R}^n$, in view of the preceding theorem, we conclude that the columns of $A \in \mathbb{R}^{n\times n}$ is linearly independent if and only if $\det(A) \neq 0$. In other words, A is invertible if and only if $\det(A) \neq 0$.

Corollary 6.6. *A matrix is invertible if and only if its determinant is nonzero.*

Determinants do not only give us a way to check if a matrix invertible, but also are able to provide an explicit formula for the inverse matrix. Let $A \in \mathbb{R}^{n\times n}$ be invertible, let $b \in \mathbb{R}^n$, and consider the equation $Ax = b$, which can be written as

$$Ax = x_1 A_1 + x_2 A_2 + \ldots + x_n A_n = b, \tag{106}$$

where $A_1, \ldots, A_n \in \mathbb{R}^n$ are the columns of A. Since A is invertible, there exists a unique solution $x = A^{-1}b \in \mathbb{R}^n$. We want to derive an explicit formula for the entries of x. To this end, invoking the properties of the determinant, we compute

$$\begin{aligned} \det(b, A_2, \ldots, A_n) &= \det(x_1 A_1 + \ldots + x_n A_n, A_2, \ldots, A_n) = x_1 \det(A_1, A_2, \ldots, A_n) \\ &= x_1 \det(A), \end{aligned} \tag{107}$$

which yields

$$x_1 = \frac{\det(b, A_2, \ldots, A_n)}{\det(A)}. \tag{108}$$

This argument can be generalized as

$$\begin{aligned} \det(\ldots, A_{i-1}, b, A_{i+1}, \ldots) &= \det(\ldots, A_{i-1}, x_1 A_1 + \ldots + x_n A_n, A_{i+1}, \ldots) \\ &= x_i \det(\ldots, A_{i-1}, A_i, A_{i+1}, \ldots) = x_i \det(A), \end{aligned} \tag{109}$$

leading to the so-called *Cramer's rule*

$$x_i = \frac{\det(\ldots, A_{i-1}, b, A_{i+1}, \ldots)}{\det(A)}, \qquad i = 1, \ldots, n. \tag{110}$$

To compute the entries of A^{-1}, we start with the observation that $AC = I$ means $AC_k = e_k$, $k = 1, \ldots, n$, where $C_k \in \mathbb{R}^n$ is the k-th column of $C = A^{-1}$, and e_k is the k-th standard basis vector in $\mathbb{R}^n$. Thus Cramer's rule yields

$$c_{ik} = \frac{\det(\ldots, A_{i-1}, e_k, A_{i+1}, \ldots)}{\det(A)}, \qquad i, k = 1, \ldots, n, \tag{111}$$

for the entries of C. We define the *adjugate of* A, as the matrix $\mathrm{adj}(A) \in \mathbb{R}^{n\times n}$ whose (i,k)-th entry is given by

$$\det(\ldots, A_{i-1}, e_k, A_{i+1}, \ldots). \tag{112}$$

Then we have

$$\det(A)A^{-1} = \mathrm{adj}(A). \tag{113}$$