

LINEAR ALGEBRA COMP GUIDE

MARIO L. GUTIERREZ ABED

DIRECT SUMS & DIRECT PRODUCTS

Definition. Let $\mathcal{F} = \{V_i \mid i \in K\}$ be any family of vector spaces over some field of scalars \mathbb{F} . The **direct product** of \mathcal{F} is the vector space

$$\prod_{i \in K} V_i = \left\{ f: K \rightarrow \bigcup_{i \in K} V_i \mid f(i) \in V_i \right\},$$

thought of as a subspace of the vector space of all functions from K to $\cup_i V_i$. ★

It will prove more useful to restrict the set of functions to those with finite support:

Definition. Let $\mathcal{F} = \{V_i \mid i \in K\}$ be a family of vector spaces over some field of scalars \mathbb{F} . The **support** of a function $f: K \rightarrow \cup_i V_i$ is the set

$$\text{supp}(f) = \{i \in K \mid f(i) \neq 0\}.$$

Thus, a function f is said to have **finite support** if $f(i) = 0$ for all but a finite number of $i \in K$. ★

Definition. The **(external) direct sum** of the family of vector spaces $\mathcal{F} = \{V_i \mid i \in K\}$ is the vector space

$$\bigoplus_{i \in K}^{\text{ext}} V_i = \left\{ f: K \rightarrow \bigcup_{i \in K} V_i \mid f(i) \in V_i, f \text{ has finite support} \right\},$$

thought of as a subspace of the vector space of all functions from K to $\cup V_i$. ★

Remark: An important special case occurs when $V_i = V$ for all $i \in K$. If we let V^K denote the set of all functions from K to V and $(V^K)_0$ denote the set of all functions in V^K that have finite support, then

$$\prod_{i \in K} V = V^K \quad \text{and} \quad \bigoplus_{i \in K}^{\text{ext}} V = (V^K)_0.$$

Note that the direct product and the external direct sum are the same for a finite family of vector spaces.

Definition. Let V be a vector space. We say that V is the **(internal) direct sum** of a family $\mathcal{F} = \{S_i \mid i \in K\}$ of subspaces of V if every vector $v \in V$ can be written in a unique way (except for order), as a finite sum of vectors from the subspaces in \mathcal{F} ; that is, if for all $v \in V$,

$$v = u_1 + \cdots + u_n \quad \text{for } u_i \in S_i.$$

Furthermore, if

$$v = w_1 + \cdots + w_m \quad \text{for } w_i \in S_i,$$

then $m = n$ and (after reindexing if necessary) we also have that $w_i = u_i$ for all $i = 1, \dots, n$.

If V is the internal direct sum of \mathcal{F} , we write

$$V = \bigoplus_{i \in K} S_i$$

and refer to each S_i as a **direct summand** of V . ★

Remark: It can be shown that the concepts of internal and external direct sums are essentially equivalent (isomorphic). For this reason, we often use the term “direct sum” to refer to either type.

Definition. If $V = S \oplus T$, then T is called a **complement** of S in V . ★

Theorem 1. A vector space V is the direct sum of a family $\mathcal{F} = \{S_i \mid i \in K\}$ of subspaces if and only if the following two conditions are met:

i) V is the sum of the S_i , i.e.

$$V = \sum_{i \in K} S_i.$$

ii) For each $i \in K$, we have

$$S_i \cap \left(\sum_{j \neq i} S_j \right) = \{0\}.$$

DUAL SPACES

Definition. Let V be a vector space over \mathbb{F} . A linear transformation $f \in \mathcal{L}(V, \mathbb{F})$ whose values lie in the base field \mathbb{F} is called a **linear functional** on V . The vector space of all linear functionals on V is denoted by V^* and is called the **algebraic dual space** of V . ★

Remark: For any $f \in V^*$, the rank-nullity theorem is

$$\dim(V) = \dim(\ker(f)) + \dim(\text{Im}(f)).$$

But since $\text{Im}(f) \subseteq \mathbb{F}$, we have either $\text{Im}(f) = 0$ (in which case f is the zero linear functional) or $\text{Im}(f) = \mathbb{F}$ (in which case f is surjective). In other words, a nonzero linear functional is surjective. Moreover, if $f \neq 0$, then

$$\text{codim}(\ker(f)) = \dim(V / \ker(f)) = 1,$$

and if $\dim(V) < \infty$ then

$$\dim(\ker(f)) = \dim(V) - \dim(V / \ker(f)) = \dim(V) - 1.$$

Thus, in dimensional terms, the kernel of a linear functional is a very “large” subspace of the domain V .

The following theorem will prove very useful:

Theorem 2. *We have the following important facts about linear functionals:*

- a) *For any nonzero vector $v \in V$, there exists a linear functional $f \in V^*$ for which $f(v) \neq 0$.*
- b) *A vector $v \in V$ is zero if and only if $f(v) = 0$ for all $f \in V^*$.*
- c) *Let $f \in V^*$. If $f(x) \neq 0$, then*

$$V = \langle x \rangle \oplus \ker(f).$$

- d) *Two nonzero linear functionals $f, g \in V^*$ have the same kernel if and only if there is a nonzero scalar λ such that $f = \lambda g$.*

Dual Basis

Let V be a vector space with basis $\mathcal{B} = \{v_i \mid i \in I\}$. For each $i \in I$, we can define a linear functional $v_i^* \in V^*$, by the orthogonality condition

$$v_i^*(v_j) = \delta_{i,j}.$$

This brings us to the following theorem:

Theorem 3. *Let V be a vector space with basis $\mathcal{B} = \{v_i \mid i \in I\}$.*

- a) *The set $\mathcal{B}^* = \{v_i^* \mid i \in I\}$ is linearly independent.*
- b) *If V is finite-dimensional, then \mathcal{B}^* is a basis for V^* , called the **dual basis** of \mathcal{B} .*

Proof of a). Notice that by applying the equation

$$0 = a_{i_1} v_{i_1}^* + \cdots + a_{i_n} v_{i_n}^*$$

to the basis vector $v_{i_k} \in \mathcal{B}$, we get

$$0 = \sum_{j=1}^k a_{i_j} v_{i_j}^*(v_{i_k}) = \sum_{j=1}^k a_{i_j} \delta_{i_j, i_k} = a_{i_k} \quad \text{for all } i_k. \quad \square$$

Proof of b). Note that for any $f \in V^*$ we have

$$\sum_j f(v_j) v_j^*(v_i) = \sum_j f(v_j) \delta_{i,j} = f(v_i),$$

and so $f = \sum_j f(v_j) v_j^*$ is in the span of \mathcal{B}^* . By part a) we already know that \mathcal{B}^* is linearly independent. Hence, \mathcal{B}^* is a basis for V^* . \square

Corollary 1. *If $\dim V < \infty$, then $\dim V^* = \dim V$.*

ADJOINT OF A LINEAR TRANSFORMATION

Theorem 4. Let V and W be finite-dimensional inner product spaces over \mathbb{F} and let $T \in \mathcal{L}(V, W)$. Then there is a unique function $T^*: W \rightarrow V$ that satisfies

$$\langle Tv, w \rangle = \langle v, T^*w \rangle \quad \forall v \in V, w \in W.$$

This function is called the **adjoint**¹ of T .

Proof. For a fixed $w \in W$, consider the function $\theta_w: V \rightarrow \mathbb{F}$ defined by

$$\theta_w(v) = \langle T(v), w \rangle.$$

It is easy to verify that θ_w is a linear functional on V and so, by the *Riesz Representation Theorem*, there exists a unique vector $x \in V$ for which

$$\theta_w(v) = \langle v, x \rangle \quad \text{for all } v \in V.$$

Hence, if $T^*(w) = x$ then

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle \quad \text{for all } v \in V.$$

This establishes the existence and uniqueness of T^* .

To show that T^* is linear, observe that

$$\begin{aligned} \langle v, T^*(\alpha w + \beta u) \rangle &= \langle T(v), \alpha w + \beta u \rangle \\ &= \alpha \langle T(v), w \rangle + \beta \langle T(v), u \rangle \\ &= \alpha \langle v, T^*(w) \rangle + \beta \langle v, T^*(u) \rangle \\ &= \langle v, \alpha T^*(w) \rangle + \langle v, \beta T^*(u) \rangle \\ &= \langle v, \alpha T^*(w) + \beta T^*(u) \rangle \end{aligned}$$

for all $v \in V$, and so

$$T^*(\alpha w + \beta u) = \alpha T^*(w) + \beta T^*(u).$$

Hence $T^* \in \mathcal{L}(W, V)$. □

Example: Let's work out an example of how the adjoint is computed.


Define $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by:

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Thus T^* will be a function from \mathbb{R}^2 to \mathbb{R}^3 . To compute T^* , fix a point $(y_1, y_2) \in \mathbb{R}^2$. Then

$$\begin{aligned} \langle (x_1, x_2, x_3), T^*(y_1, y_2) \rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2) \rangle \\ &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2) \rangle \\ &= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1) \rangle \end{aligned}$$

for all $(x_1, x_2, x_3) \in \mathbb{R}^3$.

This shows that $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$. 

¹The word *adjoint* has another meaning in linear algebra, which is related to inverses. Be warned that the two meanings for adjoint are unrelated to one another.

Here are some of the basic properties of the adjoint:

Theorem 5. *Let V and W be finite-dimensional inner product spaces. For every $\sigma, \tau \in \mathcal{L}(V, W)$, and $\alpha \in \mathbb{F}$, we have*

- $(\sigma + \tau)^* = \sigma^* + \tau^*$.
- $(\alpha\tau)^* = \bar{\alpha}\tau^*$.
- $\tau^{**} = \tau$.
- If $V = W$, then $(\sigma\tau)^* = \tau^*\sigma^*$.
- If τ is invertible, then $(\tau^{-1})^* = (\tau^*)^{-1}$.
- If $V = W$ and $p(x) \in \mathbb{R}[x]$, then $p(\tau)^* = p(\tau^*)$.

TENSOR PRODUCTS

Existence I: Intuitive but not Coordinate-Free

The universal property for bilinearity captures the essence of bilinearity *and nothing more* (as is the intent for all universal properties). To understand better how this can be done, let $\mathcal{B} = \{e_i \mid i \in I\}$ be a basis for U and let $\mathcal{C} = \{f_j \mid j \in J\}$ be a basis for V . Then a bilinear map t on $U \times V$ is uniquely determined by assigning arbitrary values to the “basis” pairs (e_i, f_j) . How can we do this and nothing more?

The answer is that we should define t on the pairs (e_i, f_j) in such a way that the images $t(e_i, f_j)$ *do not interact* and then extend by bilinearity.

In particular, for each ordered pair (e_i, f_j) , we invent a new formal symbol, say $e_i \otimes f_j$ and define T to be the vector space with basis

$$\mathcal{D} = \{e_i \otimes f_j \mid e_i \in \mathcal{B}, f_j \in \mathcal{C}\}.$$

Then define the map t by setting

$$t(e_i, f_j) = e_i \otimes f_j$$

and extending by bilinearity. This uniquely defines a bilinear map t that is as “universal” as possible among bilinear maps.

Indeed, if $g: U \times V \rightarrow W$ is bilinear, then the condition $g = \tau \circ t$ is equivalent to

$$\tau(e_i \otimes f_j) = g(e_i, f_j)$$

which uniquely defines a linear map $\tau: T \rightarrow W$. Hence, the pair (T, t) has the universal property for bilinearity.

A typical element of T is a finite linear combination

$$\sum_{i,j=1}^n \alpha_{i,j} (e_{k_i} \otimes f_{k_j}),$$

and if $u = \sum \alpha_i e_i$ and $v = \sum \beta_j f_j$, then

$$u \otimes v = t(u, v) = t\left(\sum \alpha_i e_i, \sum \beta_j f_j\right) = \sum \alpha_i \beta_j (e_i \otimes f_j).$$

Note that, as is customary, we have used the notation $u \otimes v$ for the image of *any* pair (u, v) under t . Strictly speaking, this is an abuse of the notation \otimes as we have defined it.

Confusion may arise because while the elements $u_i \otimes v_j$ form a basis for T (by definition), the larger set of elements of the form $u \otimes v$ do span T , but are definitely not linearly independent. This raises various questions, such as when a sum of the form $\sum u_i \otimes v_j$ is equal to 0, or when we can define a map τ on T by specifying the values $\tau(u \otimes v)$ arbitrarily. The first question seems more obviously challenging when we phrase it by asking when a sum of the form $\sum t(u_i, v_j)$ is 0, since there is no algebraic structure on the cartesian product $U \times V$, and so there is nothing “obvious” that we can do with this sum. The second question is not difficult to answer when we keep in mind that the set $\{u \otimes v\}$ is not linearly independent.

The notation \otimes is used in yet another way: T is generally denoted by $U \otimes V$ and called the **tensor product** of U and V . The elements of $U \otimes V$ are called **tensors** and a tensor of the form $u \otimes v$ is said to be **decomposable**:

For example, in $\mathbb{R}^2 \otimes \mathbb{R}^2$, the tensor $(1, 1) \otimes (1, 2)$ is decomposable but the tensor $(1, 1) \otimes (1, 2) + (1, 2) \otimes (2, 3)$ is not.

It is also worth emphasizing that the tensor product \otimes is not a product in the sense of a binary operation on a set, as is the case in rings and fields, for example. In fact, even when $V = U$, the tensor product $u \otimes u$ is not in U , but rather in $U \otimes U$. It is wise to remember that the decomposable tensor $u \otimes v$ is nothing more than the image of the ordered pair (u, v) under the bilinear map t , as are the basis elements $e_i \otimes f_j$.

Existence II: Coordinate-Free

The previous definition of tensor product is about as intuitive as possible, but has the disadvantage of not being coordinate free. The following customary approach to defining the tensor product does not require the choice of a basis.

Let $\mathbb{F}_{U \times V}$ be the vector space over \mathbb{F} with basis $U \times V$. Let S be the subspace of $\mathbb{F}_{U \times V}$ generated by all vectors of the form

$$\alpha(u, w) + \beta(v, w) - (\alpha u + \beta v, w)$$

and

$$\alpha(u, v) + \beta(u, w) - (u, \alpha v + \beta w),$$

where $\alpha, \beta \in \mathbb{F}$ and u, v , and w are in the appropriate spaces. Note that these vectors are 0 if we replace the ordered pairs by tensors according to our previous definition.

The quotient space

$$U \otimes V = (\mathbb{F}_{U \times V})/S$$

is also called the **tensor product** of U and V . The elements of $U \otimes V$ have the form

$$\left(\sum \alpha_i (u_i, v_i)\right) + S = \sum \alpha_i [(u_i, v_i) + S].$$

However, since by definition $\alpha(u, v) - (\alpha u, v) \in S$ and $\alpha(u, v) - (u, \alpha v) \in S$, we can “absorb” the scalar in either coordinate, that is,

$$\alpha[(u, v) + S] = (\alpha u, v) + S = (u, \alpha v) + S,$$

and so the elements of $U \otimes V$ can be written simply as

$$\sum [(u_i, v_i) + S].$$

It is customary to denote the coset $(u, v) + S$ by $u \otimes v$ and so any element of $U \otimes V$ has the form

$$\boxed{\sum u_i \otimes v_i}$$

as in the previous “non-coordinate-free” definition.

Finally, the map $t: U \times V \rightarrow U \otimes V$ is defined by

$$t(u, v) = u \otimes v.$$

Theorem 6. *Let U and V be vector spaces. Then the pair $(U \otimes V, t: U \times V \rightarrow U \otimes V)$ has the universal property for bilinearity, as measured by linearity.*

Proposition 1 (A Basis for the Tensor Product Space). *Suppose V_1, \dots, V_k are real vector spaces of dimensions n_1, \dots, n_k , respectively. For each $j = 1, \dots, k$, suppose $(E_1^{(j)}, \dots, E_{n_j}^{(j)})$ is a basis for V_j . Then the set*

$$\mathcal{C} = \left\{ E_{i_1}^{(1)} \otimes \dots \otimes E_{i_k}^{(k)} \mid 1 \leq i_1 \leq n_1, \dots, 1 \leq i_k \leq n_k \right\}$$

is a basis for $V_1 \otimes \dots \otimes V_k$, which therefore has dimension equal to $n_1 \dots n_k$.

Aside

Definition. *An algebra A over a field \mathbb{F} is said to be a **graded algebra** if, as a vector space over \mathbb{F} , A can be written in the form*

$$A = \bigoplus_{i=0}^{\infty} A_i$$

for subspaces A_i of A , and where multiplication behaves nicely in the sense that

$$A_i A_j \subseteq A_{i+j}.$$

*The elements of A_i are said to be **homogeneous of degree i** . If $a \in A$ is written*

$$a = a_{i_1} + \dots + a_{i_n} \quad \text{for } a_{i_k} \in A_{i_k}, \text{ and } i_k \neq i_j,$$

*then a_{i_k} is called the **homogeneous component** of a of degree i .*

★

Definition. *A **graded ideal** I in a graded algebra $A = \bigoplus A_i$ is an ideal I for which, as a subspace of A , we have that $I = \bigoplus_{i=0}^{\infty} (I \cap A_i)$.*

★

Remark: Note that $I \cap A_i$ is not, in general, an ideal. For example, the ideal I of $\mathbb{F}[x]$ consisting of all polynomials with zero constant term is graded. However, the ideal

$$J = \langle 1 + x \rangle = \{p(x)(1 + x) \mid p(x) \in \mathbb{F}[x]\}$$

generated by $1 + x$ is not graded, since $\mathbb{F}_i[x]$ contains only monomials and so $J \cap \mathbb{F}_i[x] = \{0\}$.

Theorem. *Let A be a graded algebra. An ideal I of A is graded if and only if it is generated by homogeneous elements of A .*

EXTERIOR (GRASSMANN) PRODUCTS

Let S_p be the symmetric group on $\{1, \dots, p\}$. For each $\sigma \in S_p$, the multilinear map $f_\sigma: V^{\times p} \rightarrow T^p(V)$ defined by

$$f_\sigma(v_1, \dots, v_p) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)}$$

determines (by universality) a unique linear operator λ_σ on $T^p(V)$ for which

$$\lambda_\sigma(v_1 \otimes \cdots \otimes v_p) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)}.$$

Let $\mathcal{B} = \{e_1, \dots, e_n\}$ be a basis for V . Since the set

$$\mathcal{C} = \{e_{i_1} \otimes \cdots \otimes e_{i_p} \mid e_{i_j} \in \mathcal{B}\}$$

is a basis for $T^p(V)$ and λ_σ is a bijection of \mathcal{C} , it follows that λ_σ is an automorphism of $T^p(V)$. A tensor $t \in T^p(V)$ is said to be **symmetric** if $\lambda_\sigma(t) = t$ or **antisymmetric** (also, **alternating**) if $\lambda_\sigma(t) = \text{sgn}(\sigma)t$ for all permutations $\sigma \in S_p$.

Now let

$$\begin{aligned} I_p &= \langle \text{sgn}(\sigma)\lambda_\sigma(t) - t \mid t \in T^p(V), \sigma \in S_p \rangle \\ &= \langle v_1 \otimes \cdots \otimes v_p \mid v_i = v_j \text{ for some } i \neq j \rangle. \end{aligned}$$

Then we have the quotient space

$$\frac{T^p(V)}{I_p} = \Lambda^p(V),$$

which we refer to as the **antisymmetric** (also, **alternating**) **tensor space** of degree p of V or the **exterior algebra** of degree p of V . The product defined on this algebra (or tensor space) is denoted by “ \wedge ” and it is called the **wedge product** or **exterior product**.

DETERMINANTS

The universal property for antisymmetric multilinear maps provides us with the following proposition:

Proposition. *Let V be a vector space of dimension n over a field \mathbb{F} . Let $E = (e_1, \dots, e_n)$ be an ordered basis for V . Then there is a unique antisymmetric n -linear form $d: V^{\times n} \rightarrow \mathbb{F}$ for which*

$$d(e_1, \dots, e_n) = 1.$$

Proof. According to the universal property for antisymmetric n -linear forms, for every such form $f: V^{\times n} \rightarrow \mathbb{F}$, there is a unique linear map $\tau_f: \Lambda^n V \rightarrow \mathbb{F}$ for which

$$\tau_f(e_1 \wedge \cdots \wedge e_n) = f(e_1, \dots, e_n) = 1.$$

But the dimension of $\Lambda^n V$ is $\binom{n}{n} = 1$ and $\{e_1 \wedge \cdots \wedge e_n\}$ is a basis for $\Lambda^n V$. Hence, there is only one linear map $\sigma: \Lambda^n V \rightarrow \mathbb{F}$ with $\sigma(e_1 \wedge \cdots \wedge e_n) = 1$. It follows that if f and g are two such forms, then

$$f(e_1, \dots, e_n) = \sigma(e_1 \wedge \cdots \wedge e_n) = g(e_1, \dots, e_n)$$

and the antisymmetry of f and g imply that f and g agree on every permutation of (e_1, \dots, e_n) . Since f and g are multilinear, we must have $f = g$. \square

We wish to construct the unique antisymmetric form d (which we will call the *determinant*) guaranteed by the previous result. In other words, given a free module V of rank n , we will define its **determinant** to be

$$\det V = \Lambda^{\max} V = \Lambda^n V.$$

Let V be a vector space. Then, given the ordered basis $E = (e_1, \dots, e_n)$, we can view V as the space \mathbb{F}^n of coordinate vectors and view $V^{\times n}$ as the space $\text{Mat}_n(\mathbb{F})$ of $n \times n$ matrices, via the isomorphism

$$(v_1, \dots, v_n) \mapsto \begin{pmatrix} [v_1]_1 & \cdots & [v_n]_1 \\ \vdots & & \vdots \\ [v_1]_n & \cdots & [v_n]_n \end{pmatrix},$$

where all coordinate matrices are with respect to the basis E .

From this viewpoint, d becomes an antisymmetric n -form on the columns of a matrix $A = (a_{i,j})$ given by

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

which we call the **determinant** of the matrix A .

Theorems to Prove

Theorem. *Prove that if $\dim V = n$ and $T \in \mathcal{L}(V)$, then $\det T \neq 0$ if and only if T is invertible.*

Proof. Note that $\det T \neq 0$ if and only if, for some basis e_1, \dots, e_n , we have that $(Te_1) \wedge \cdots \wedge (Te_n)$ is nonzero, which occurs if and only if the list (Te_1, \dots, Te_n) is linearly independent. This occurs if and only if T is an isomorphism, i.e. T is invertible. \square

Theorem. *If $A, B \in \text{Mat}_n(\mathbb{F})$, then $\det(AB) = \det A \det B$.*

Proof. Consider the map $f_A: \text{Mat}_n(\mathbb{F}) \rightarrow \mathbb{F}$ defined by

$$f_A(X) = \det(AX).$$

We can consider f_A as a function on the columns of X and write

$$f_A: (X^{(1)}, \dots, X^{(n)}) \mapsto (AX^{(1)}, \dots, AX^{(n)}) \mapsto \det(AX).$$

Now, this map is multilinear since multiplication by A is distributive and the determinant is multilinear. For example, let $y \in \mathbb{F}^n$ and let X' come from X by replacing the first column by y . Then

$$\begin{aligned} f_A(\alpha X^{(1)} + \beta y, \dots, X^{(n)}) &\mapsto (\alpha AX^{(1)} + \beta Ay, \dots, AX^{(n)}) \\ &\mapsto \alpha \det(AX) + \beta \det(AX') \\ &= \alpha f_A(X) + \beta f_A(X'). \end{aligned}$$

The map f_A is also alternating since \det is alternating and interchanging two coordinates in $(X^{(1)}, \dots, X^{(n)})$ is equivalent to interchanging the corresponding columns of AX .

Thus, f_A is an antisymmetric n -linear form and so it must be a scalar multiple of the determinant function, say $f_A(X) = \gamma \det(X)$. Then

$$\det(AX) = f_A(X) = \gamma \det(X).$$

Setting $X = I_n$ gives $\det A = \gamma$ and so

$$\det(AX) = \det A \det X,$$

as desired. □

Theorem. *Every vector space has a basis.*

Proof. The idea is that a basis can be constructed as a maximal linearly independent set, and this maximal set will be found by using Zorn's lemma.²

Let V be a nonzero vector space and let S be the set of linearly independent sets in V . Since a single nonzero $v \in V$ is a linearly independent set, we have that $\{v\} \in S$, which indicates that S is nonempty.

Now for two linearly independent sets L and L' in V , we declare that $L \leq L'$ if $L \subset L'$, where \leq represents the partial ordering on S by inclusion. It is easy to see that any subset of a linearly independent set is also a linearly independent set, so if $L \in S$, then any subset of L is also in S .

Now that we have defined a partial order on S , let $\{L_\lambda\}_{\lambda \in \Lambda}$ be a totally ordered subset of S , i.e. a chain on S . That is, every L_λ is a linearly independent set in V and for any L_α and L_β in our chain we have $L_\alpha \subset L_\beta$ or $L_\beta \subset L_\alpha$. An upper bound for the L_λ 's in S is the union

$$(\clubsuit) \quad L = \bigcup_{\lambda \in \Lambda} L_\lambda.$$

The next step is to check whether L is indeed a linearly independent set, so that L is an element of S ; once that is settled then L would be an upper bound in S since $L_\lambda \subset L \ \forall \lambda \in \Lambda$.

Let us take any finite set of vectors $v_1, \dots, v_n \in L$, and show that they are linearly independent. Each v_k is in some L_λ , say $v_1 \in L_{\lambda_1}, \dots, v_n \in L_{\lambda_n}$. Since the L_λ 's are totally ordered, one of the sets $L_{\lambda_1}, \dots, L_{\lambda_n}$ contains the others. That means v_1, \dots, v_n are all in a common L_λ , so they are linearly independent, as desired.

Zorn's lemma now tells us that S contains a maximal element: there is a linearly independent set $L \in V$ that is not contained in any larger linearly independent set in V . We will show that L spans V , so it is a basis.

²Here's Zorn's Lemma, for reference:

Zorn's Lemma: If P is a partially ordered set in which every chain has an upper bound, then P has a maximal element.

If $\text{span}(L)$ does not span V , then $\text{span}(L) \neq V$, so we can pick $v \in V$ with $v \notin \text{span}(L)$. Then L is a proper subset of $L \cup \{v\}$. We will show $L \cup \{v\}$ is linearly independent, which contradicts the maximality of L and thus proves that $\text{span}(L) = V$.

To prove that $L \cup \{v\}$ is linearly independent, assume otherwise. That is, take a sum

$$\sum_{i=1}^k c_i v_i = 0$$

where the c_i 's are not all 0 and the v_i 's are taken from $L \cup \{v\}$. Since the elements of L are linearly independent, one of the v_i 's with a nonzero coefficient must be v . We can re-index and suppose $v_k = v$, so $c_k \neq 0$. Then we must have $k \geq 2$, since otherwise $c_1 v = 0$, which is impossible since $v \neq 0$ and the coefficient of v is nonzero. Consequently we have

$$\begin{aligned} 0 &= c_k v + \sum_{i=1}^{k-1} c_i v_i \\ (\dagger) \quad &\implies c_k v = - \sum_{i=1}^{k-1} c_i v_i. \end{aligned}$$

Multiplying both sides of (2) by $1/c_k$, we get

$$v = \sum_{i=1}^{k-1} \left(-\frac{c_i}{c_k} \right) v_i,$$

which shows that $v \in \text{span}(L)$. ($\Rightarrow \Leftarrow$)

This is the contradiction we wanted because by assumption $v \notin \text{span}(L)$. Hence $L \cup \{v\}$ is a linearly independent set, and we are done. \square

Theorem (A closed subspace of a Hilbert space has an orthogonal complement). *Suppose S is a closed subspace of a Hilbert space \mathcal{H} and $f \in \mathcal{H}$. Then*

- *There exists a (unique) element $g_0 \in S$ which is closest to f , in the sense that*

$$\|f - g_0\| = \inf_{g \in S} \|f - g\|.$$

- *The element $f - g_0$ is perpendicular to S , that is,*

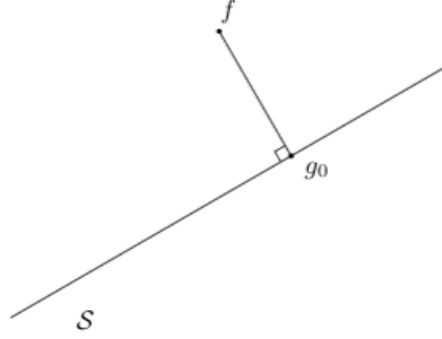
$$\langle f - g_0, g \rangle = 0 \quad \forall g \in S.$$

The situation described in the theorem can be visualized in Figure 1 below:

Proof. If $f \in S$, then we choose $f = g_0$, and there is nothing left to prove. Otherwise, we let $d = \inf_{g \in S} \|f - g\|$, and note that we must have $d > 0$ since $f \notin S$ and S is closed. Consider a sequence $\{g_n\}_{n=1}^{\infty}$ in S such that

$$\|f - g_n\| \rightarrow d \quad \text{as } n \rightarrow \infty.$$

We claim that $\{g_n\}$ is a Cauchy sequence whose limit will be the desired element g_0 . In fact, it would suffice to show that a subsequence of $\{g_n\}$ converges, and this is immediate in the finite-dimensional case because a closed ball is compact. However, in general this compactness fails, and so a more intricate argument is needed at this point.


 FIGURE 1. Nearest element to f in S .

To prove our claim, we use the *parallelogram law*, which states that in a Hilbert space \mathcal{H} , we have

$$(1) \quad \|A + B\|^2 + \|A - B\|^2 = 2[\|A\|^2 + \|B\|^2] \quad \text{for all } A, B \in \mathcal{H}.$$

The simple verification of this equality, which consists of writing each norm in terms of the inner product, is left to the reader. Putting $A = f - g_n$ and $B = f - g_m$ in the parallelogram law, we find

$$\|2f - (g_n + g_m)\|^2 + \|g_m - g_n\|^2 = 2[\|f - g_n\|^2 + \|f - g_m\|^2].$$

However S is a subspace, so the quantity $1/2(g_n + g_m)$ belongs to S ; hence

$$\|2f - (g_n + g_m)\| = 2\|f - \frac{1}{2}(g_n + g_m)\| \geq 2d.$$

Therefore,

$$\begin{aligned} \|g_m - g_n\|^2 &= 2[\|f - g_n\|^2 + \|f - g_m\|^2] - \|2f - (g_n + g_m)\|^2 \\ &\leq 2[\|f - g_n\|^2 + \|f - g_m\|^2] - 4d^2. \end{aligned}$$

By construction, we know that $\|f - g_n\| \rightarrow d$ and $\|f - g_m\| \rightarrow d$ as $n, m \rightarrow \infty$, so the above inequality implies that $\{g_n\}$ is a Cauchy sequence. Since \mathcal{H} is complete and S closed, the sequence $\{g_n\}$ must have a limit g_0 in S , and then it satisfies $d = \|f - g_0\|$.

We prove that if $g \in S$, then $g \perp (f - g_0)$. For each ε (positive or negative), consider the perturbation of g_0 defined by $g_0 - \varepsilon g$. This element belongs to S , so

$$\|f - (g_0 - \varepsilon g)\|^2 \geq \|f - g_0\|^2.$$

Since $\|f - (g_0 - \varepsilon g)\|^2 = \|f - g_0\|^2 + \varepsilon^2\|g\|^2 + 2\varepsilon\text{Re}(f - g_0, g)$, we find that

$$(2) \quad 2\varepsilon\text{Re}(f - g_0, g) + \varepsilon^2\|g\|^2 \geq 0.$$

If $\text{Re}(f - g_0, g) < 0$, then taking ε small and positive contradicts (2). If $\text{Re}(f - g_0, g) > 0$, a contradiction also follows by taking ε small and negative. Thus we must have $\text{Re}(f - g_0, g) = 0$. By considering the perturbation $g_0 - i\varepsilon g$, a similar argument gives $\text{Im}(f - g_0, g) = 0$, and hence $(f - g_0, g) = 0$.

Finally, the uniqueness of g_0 follows from the above observation about orthogonality. Suppose \tilde{g}_0 is another point in S that minimizes the distance to f . By taking $g = g_0 - \tilde{g}_0$ in our last argument we find $(f - g_0) \perp (g_0 - \tilde{g}_0)$, and the Pythagorean theorem gives

$$\|f - \tilde{g}_0\|^2 = \|f - g_0\|^2 + \|g_0 - \tilde{g}_0\|^2.$$

Since by assumption $\|f - \tilde{g}_0\|^2 = \|f - g_0\|^2$, we conclude that $\|g_0 - \tilde{g}_0\| = 0$, as desired. \square