Definition. A **vector space** V over a **field** F is a set with two operations:
(**vector addition**) For any x and y in V, there is a uniquely defined element $x+y$ in V.
(**scalar multiplication**) For any a in F and x in V, there is a uniquely defined element $a \cdot x$ in V.
Satisfying the following eight conditions (VS1) – (VS 8):
(VS1) $x+y = y+x$ for any x, y in V. (commutativity of addition)
(VS2) $(x+y)+z = x+(y+z)$ for any x,y,z in V (associativity of addition)
(VS3) There is an element in V denoted by 0 such that $x+0 = x$ for all x in V.
(VS4) For each x in V there is some y in V such that $x+y=0$.
(VS5) $1 \cdot x = x$ for all x in V (where 1 is the multiplicative identity of F).
(VS6) $a \cdot (b \cdot x) = (a \cdot b) \cdot x$ for all x in V and a, b in F. (associativity of scalar multiplication)
(VS7) $a \cdot (x+y) = a \cdot x + a \cdot y$ for all a in F and x, y in V
(VS8) $(a+b) \cdot x = a \cdot x + b \cdot x$ for all a, b in F and x in V.
} distributive laws

- Elements of V are called **vectors**, and elements of F are called **scalars**.
- Usually F will be either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$.

Example 1. Given a field F, consider the set

$$F^n = \{ (x_1, x_2, \dots, x_n) : x_i \in F \} \quad \text{(the set of all } n\text{-tuples of elements from } F).$$

We define the operations of vector addition and scalar multiplication in the following way:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1+y_1, \dots, x_n+y_n)$$

$$a \cdot (x_1, \dots, x_n) = (a \cdot x_1, \dots, a \cdot x_n),$$

where $(x_1, \dots, x_n), (y_1, \dots, y_n)$ are arbitrary elements of F^n and a is an arbitrary element of F.
(and $x_i + y_i$ and $a \cdot x_i$ are calculated in F).
With these operations F^n is a vector space over the field F. (One has to check that (VS1) – (VS8) hold. Do it!)
In particular, if $n=2$ and $F = \mathbb{R}$, we obtain the familiar space $\mathbb{R}^2$ of vectors on the plane.

Example 2. Let F be a field, and let **$P(F)$** denote the set of all polynomials with coefficients in F. That is, $P(F)$ consists of all expressions of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

for some $n \geq 0$, with a_i in F for all $i = 0, 1, \dots, n$. If $a_i = 0$ for all $i = 0, 1, \dots, n$ then $p(x)$ is called the **zero polynomial**.
The **degree** of a non-zero polynomial is the largest i such that $a_i \neq 0$. (the degree of the zero polynomial is defined to be -1)
Two polynomials $p(x) = a_n x^n + \dots + a_1 x + a_0$ and $q(x) = b_m x^m + \dots + b_1 x + b_0$ are equal if $m=n$ and $a_i = b_i$ for all $i = 0, 1, \dots, n$.
We define addition and scalar multiplication on $P(F)$ in the usual way:
Addition: Given $p(x), q(x)$ as above, suppose that $m \leq n$. Then we can also write $q(x)$ as $b_n x^n + \dots + b_1 x + b_0$, where $b_{m+1} = b_{m+2} = \dots = b_n = 0$
Then we define $p(x) + q(x) = (a_n + b_n) x^n + (a_{n-1} + b_{n-1}) x^{n-1} + \dots + (a_1 + b_1) x + (a_0 + b_0)$.
Scalar multiplication: For any c in F, define $c \cdot p(x) = c a_n x^n + c a_{n-1} x^{n-1} + \dots + c a_1 x + c a_0$.
(Equivalently, $p+q$ can be defined as the polynomial satisfying $(p+q)(x) = p(x) + q(x)$ for all x in F, and cp as the polynomial satisfying $(cp)(x) = c \cdot p(x)$ for all x in F).
With these operations, $P(F)$ is a vector space over F. (Again, need to check that (VS1) – (VS8) hold.)

Example 3. Let $M_2(\mathbb{R})$ denote the set of all 2×2 matrices with entries from $\mathbb{R}$.
We define addition and scalar multiplication in the familiar way:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} \quad \text{and} \quad \alpha \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{pmatrix}$$

for any $a, b, c, d, e, f, g, h, \alpha$ in $\mathbb{R}$.
With these operations, $M_2(\mathbb{R})$ is a vector space over $\mathbb{R}$.
Analogously, the space $M_{m \times n}(\mathbb{R})$ of all $m \times n$ matrices is a vector space.

Example 4 The most boring vector space, aka **the zero vector space**.
- It consists of a single vector 0 (so $V = \{0\}$).

Operations of addition and scalar multiplication are defined by:

$0 + 0 = 0$

$\alpha \cdot 0 = 0$ for all α in F.
With these operations, V is a vector space over F. (check it!)

Basic properties of vector spaces

Now, we will see some of the basic properties of vector spaces (we will deduce them as logical consequences of the axioms (VS1)–(VS8).)

Theorem 1.1 (Cancellation law)
Let V be a vector space, and x, y, z be arbitrary elements of V.
If $x+z = y+z$, then $x=y$.

Proof. As V is a vector space, it satisfies all of the properties (VS1)–(VS8).
By (VS4) there exists an element $\tilde{z}$ in V such that $z+\tilde{z}=0$. We have:

$$x \underset{\text{by (VS3)}}{=} x+0 \underset{\text{by the choice of } \tilde{z}}{=} x+(z+\tilde{z}) \underset{\text{by (VS2)}}{=} (x+z)+\tilde{z} \underset{\text{by assumption}}{=} (y+z)+\tilde{z} \underset{\text{by (VS2) again}}{=} y+(z+\tilde{z}) \underset{\text{by the choice of } \tilde{z}}{=} y+0 \underset{\text{by (VS3) again}}{=} y.$$

Thus $x=y$.

By commutativity of addition (VS1) we also have: if $z+x = z+y$, then $x=y$.

Corollary 1. The vector 0 described in (VS3) is unique. (and is called the zero vector of V).

Proof. Suppose that 0 and $0'$ are two elements in a vector space V that both satisfy (VS3).
Then for any x in V we have:

$$x+0 \underset{\text{by (VS3) for } 0}{=} x \underset{\text{by (VS3) for } 0'}{=} x+0'$$

By the cancellation law it follows that $0=0'$.

For example, in F^n the zero vector is $\overbrace{(0,0,\dots,0)}^{n \text{ times}}$, and in $M_2(\mathbb{R})$ the zero vector is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Corollary 2. For any x in V, the vector y described in (VS4) is unique.
(it is called the additive inverse of x and is denoted by $-x$).

Proof. Let x in V be arbitrary, and suppose that y_1 and y_2 are two elements in V both satisfying (VS4).
That is, $x+y_1 = 0 = x+y_2$.
By cancellation law this implies $y_1 = y_2$.

Finally, we state some further useful properties of vector spaces.

Theorem 1.2. Let V be a vector space over a field F.
For all x in V and a in F we have:

1) $\underset{\text{scalar in } F}{0} \cdot x = \underset{\text{the zero vector of } V}{0}$

2) $(-a)\cdot x = \underbrace{-(ax)}_{\text{the additive inverse of the vector } ax} = a\cdot(-x)$.

3) $a \cdot 0 = 0$ — the zero vector in V

Proof. Problem Set 1.

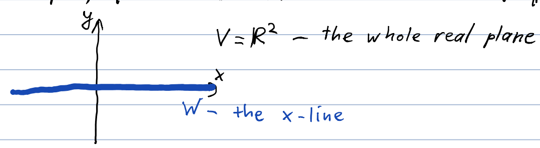
Subspaces

Definition. Let V be a vector space over a field F.
A subset W of V ($W \subseteq V$) is a subspace of V if W itself is a vector space over F, with respect to the addition and scalar multiplication defined on V.

That is, W satisfies all of the properties (VS1)–(VS8).

Example 1 Let $n \geq 1$ be an integer and F a field.
Recall that the vector space F^n is the set $\{(x_1, \dots, x_n) : x_i \in F\}$ with addition and scalar multiplication

given by $(x_1,\dots,x_n)+(y_1,\dots,y_n)=(x_1+y_1,\dots,x_n+y_n)$ and $a\cdot(x_1,\dots,x_n)=(ax_1,\dots,ax_n)$.
Consider the subset
$W=\{(x_1,\dots,x_{n-1},0) : x_i\in F\}$.
Then W is a subspace of F^n (we will prove it later).
For example, for $F=\mathbb{R}$ and $n=2$ we have $V=\mathbb{R}^2=\{(x_1,x_2) : x_1,x_2\in\mathbb{R}\}$, $W=\{(x_1,0): x_1\in\mathbb{R}\}$:



Example 2. Let F be a field, and recall that $P(F)$ is the vector space of all polynomials with coefficients from F.
For an integer $n\geq 0$, consider the subset $P_n(F)\subseteq P(F)$ consisting of all polynomials of degree $\leq n$.
Then $P_n(F)$ is a subspace of $P(F)$.

Example 3. For any vector space V, V itself and $\{0\}$ are both subspaces of V.

Example 4. Let S be a non-empty set and F a field.
Let $\mathcal{F}(S,F)$ denote the set of all maps from S to F.
(A map $f: S\to F$ takes x in S as an input and returns $f(x)$ in F as an output.
Synonym: function).
Two maps f and g in $\mathcal{F}(S,F)$ are equal if $f(x)=g(x)$ for all x in S.
For any f and g in $\mathcal{F}(S,F)$ and c in F we define $f+g$ and $c\cdot f$ in $\mathcal{F}(S,F)$ by taking
$(f+g)(x)=f(x)+g(x)$ for each x in S (where "+" and "$\cdot$" on the right side are calculated in F).
$(c\cdot f)(x) = c\cdot f(x)$
With these operations $\mathcal{F}(S,F)$ is a vector space (check!)
In particular, $\mathcal{F}(\mathbb{R},\mathbb{R})$ consists of all real-valued functions defined on real numbers.
Let $C(\mathbb{R})$ denote the set of all continuous real functions.
Then $C(\mathbb{R})$ is a subset of $\mathcal{F}(\mathbb{R},\mathbb{R})$, and we will see that it is a subspace as well.

Let now V be a vector space, and let $W\subseteq V$ be a subset of V. According to the definition, in order to show that W is a subspace, we have to check that all of the properties (VS1)–(VS8) hold in W, with respect to addition and scalar multiplication defined on V.
Turns out that we actually need to check less in this situation.

Theorem 1.3. Let V be a vector space over F, and let $W\subseteq V$ be a subset of V. (need not be a subspace!)
Then W is a subspace of V if and only if the following holds for W:
(a) $0\in W$ (that is, the zero vector of V is in W).
(b) If $x,y\in W$ then $x+y\in W$ (that is, W is closed under addition)
(c) If $x\in W$ and $c\in F$ then $c\cdot x\in W$ (that is, W is closed under scalar multiplication)

Proof.
1) First we assume that W is a subspace of V and show that (a), (b), (c) hol
"W is a subspace of V" means that it is a vector space under the operations of addition and scalar multiplication defined on V.
In particular:
• for any x,y in W and c in F, $x+y$ and $c\cdot x$ must also be in W. So (b) and (c) hold.
• (VS3) holds in W, that is there is some $0'\in W$ such that $x+0'=x$ for all $x\in W$.
In particular, $0'+0'=0'$.
But as $0'\in V$, also $0'+0=0'$ (since $0\in V$ satisfies (VS3) in V.)
By cancellation law in the vector space V, this implies that $0'=0$. In particular, $0\in W$ and (a) holds.
2) Conversely, suppose (a), (b), (c) holds, and we will show that then W is not just a subset of V, but also a subspace. For this we have to check that W satisfies (VS1)–(VS8).
• As V is a vector space, (VS1), (VS2), (VS5), (VS6), (VS7), (VS8) hold for all elements of V. As W is a subset of V,

they hold in particular for all elements of W. (so we get all these properties in W for free).

- Remains to check that (VS3) and (VS4) hold in W.

  (VS3): holds by (a).

  (VS4): Let $x \in W$ be arbitrary. We need to find some $y \in W$ such that $x+y=0$. We know that in V there is such an element, namely the additive inverse $-x$. We show that actually $-x \in W$ as wel

  As -1 is a scalar in F, by (c) we have $(-1)\cdot x = W$.

  As V is a vector space, it satisfies $-x = (-1)\cdot x$, by Theorem 1.2.

  Thus $-x \in W$, as wanted.

Now we return to the examples and verify our claims using Theorem 1.3.

Example 1. $V = F^n$, $W = \{(x_1, \ldots, x_{n-1}, 0) : x_i \in F\} \leq V$.

We check that W satisfies (a),(b),(c), and so indeed W is a subspace of V by Theorem 1.3.

(a) The zero vector of V is $\overbrace{(0,0,\ldots,0)}^{n}$, it is in W (taking $x_1 = \ldots = x_{n-1} = 0$).

(b),(c) W is closed under addition and scalar multiplication.

Given any $(x_1, \ldots, x_{n-1}, 0), (y_1, \ldots, y_{n-1}, 0)$ in W and a in F we have:

$(x_1, \ldots, x_{n-1}, 0) + (y_1, \ldots, y_{n-1}, 0) = (x_1+y_1, \ldots, x_{n-1}+y_{n-1}, \underbrace{0+0}_{=0})$ — also in W.

$a\cdot(x_1, \ldots, x_{n-1}, 0) = (a\cdot x_1, \ldots, a\cdot x_{n-1}, \underbrace{a\cdot 0}_{=0})$

Similarly, $W = \{(0, x_1, \ldots, x_{n-1}) : x_i \in F\}$ is also a subspace of F^n.

Example 2. $V = P(F)$, $W = P_n(F)$ — the set of all polynomials of degree at most n.

(a) The zero vector in $P(F)$ is the zero polynomial $p_0(x) = a_n x^n + \ldots + a_1 x + a_0$ with $a_n = \ldots = a_1 = a_0 = 0$. Its degree is by definition -1, so p_0 is in $P_n(F)$.

(b),(c) If both $p(x)$ and $q(x)$ are in $P_n(F)$, that is they have degree at most n, then $p(x)+q(x)$ also has degree at most n, and thus $p+q$ is in $P_n(F)$ as well.

If $p(x)$ has degree $\leq n$ and $a \in F$, then $a\cdot p(x)$ has degree $\leq n$ as well, and so $a.p \in P_n(F)$.

Example 4. $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ – all real-valued functions, $W = C(\mathbb{R})$ — continuous real-valued functions.

The zero vector in $\mathcal{F}$ is the function given by $f(x) = 0$ for all x in $\mathbb{R}$.

By basic calculus we know that all constant functions are continuous, that sum of any two continuous functions is continuous, and that a product of a constant function and a continuous function is also a continuous.

Thus $C(\mathbb{R})$ satisfies the conditions (a),(b),(c) in Theorem 1.3.

We can form new subspaces from the old ones.

Theorem 1.4. Let V be a vector space over F.

If $W_1, \ldots, W_n$ are subspaces of V, then the set $W = W_1 \cap W_2 \cap \ldots \cap W_n$ is also a subspace of V.

Proof. We check that W satisfies (a),(b),(c) and apply Theorem 1.3.

By assumption each of W_i, $i = 1, \ldots, n$, is a subspace, and so satisfies (a),(b),(c).

(a) As each of W_i satisfies (a), we have $0 \in W_i$ for all $i = 1, \ldots, n$.

But this means that $0 \in W_1 \cap \ldots \cap W_n = W$ as well.

(b) Let $x, y \in W$, which means precisely that $x, y \in W_i$ for all $i = 1, \ldots, n$.

As each of W_i satisfies (b), $x+y \in W_i$ for each $i = 1, \ldots, n$.

Hence $x+y \in W_1 \cap \ldots \cap W_n$.

(c) If $x \in W$, then $x \in W_i$ for each $i = 1, \ldots, n$.

Then for any $c \in F$, $c.x \in W_i$ for each $i = 1, \ldots, n$ (as W_i satisfies (c)).

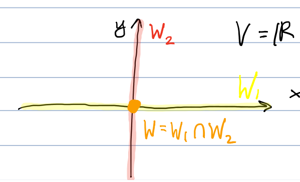
Hence $c\cdot x \in W_1 \cap \ldots \cap W_n = W$.

Example. Let $V = \mathbb{R}^2$.

Let $W_1 = \{(x_1, 0) : x_1 \in \mathbb{R}\}$ and let $W_2 = \{(0, x_2) : x_2 \in \mathbb{R}\}$.

We already know that both W_1 and W_2 are subspaces of V.

Then $W_1 \cap W_2 = \{(0,0)\}$ — the zero subspace of $\mathbb{R}^2$.



However, the union $W = W_1 \cup W_2$ of two subspaces W_1, W_2 of V need not be a subspace of V in general!
(see Problem Set 1).

Linear Combinations and Span.

Definition. Let V be a vector space and $S \subseteq V$ a non-empty subset of V.
A vector v in V is called a linear combination of vectors of S if there exist a finite number of elements $u_1, u_2, \ldots, u_n$ in S and scalars $a_1, \ldots, a_n$ in F such that $v = a_1 u_1 + a_2 u_2 + \ldots + a_n u_n$.
In this case we also say that v is a linear combination of $u_1, \ldots, u_n$ and call $a_1, \ldots, a_n$ the coefficients of the linear combination.

- In any vector space V, we always have $0v = 0$ for each $v \in V$. Thus the zero vector is a linear combination of any non-empty subset of V.
- Given $v \in V$ and $u_1, \ldots, u_n \in V$, how can one determine whether v is a linear combination of the vectors $u_1, \ldots, u_n$?

That is, we need to understand if it is possible to find scalars $a_1, \ldots, a_n \in F$ such that $a_1 u_1 + \ldots + a_n u_n = v$.
This question often reduces to solving a system of linear equations.

Example. Let $V = \mathbb{R}^2$, let $v = (1,5)$ and $u_1 = (2,0)$, $u_2 = (3,-1)$.
We must determine whether there are scalars $a_1, a_2 \in \mathbb{R}$ such that $v = a_1 u_1 + a_2 u_2$, or:
$(1,5) = a_1(2,0) + a_2(3,-1) = (2a_1, 0) + (3a_2, -a_2) = (2a_1 + 3a_2, -a_2)$.
So v is a linear combination of u_1, u_2 iff there are real numbers $a_1, a_2 \in \mathbb{R}$ such that the system of linear equations

(1) $1 = 2a_1 + 3a_2$
(2) $5 = -a_2$

is satisfied.
Note that then necessarily $a_2 = -5$ by (2), hence $1 = 2a_1 + 3(-5)$, so $2a_1 = 16$, so $a_1 = 8$.
Then $(a_1, a_2) = (8, -5)$ is a solution, showing that indeed v is a linear combination of u_1, u_2.
(For a more involved example see Textbook, Section 1.4 and Problem Set 2).

- Theorem 1.3 allows us to determine when a subset S of a vector space V is already a subspace.

Now we discuss how, starting with an arbitrary subset $S \subseteq V$ (which may not be a subspace of V itself), to find a subspace of V "generated" by it.

Definition. Let S be any subset of a vector space V.
The span of S, denoted $\mathrm{Span}(S)$, is the set consisting of all linear combinations of the vectors in S. That is,
$\mathrm{Span}(S) = \{a_1 u_1 + \ldots + a_n u_n : n \in \mathbb{N}, a_i \in F, u_i \in S\} \subseteq V$.

As the empty set $\emptyset$ is a subset of any vector space V, $\mathrm{Span}(\emptyset)$ also needs to be defined.
For convenience, we define $\mathrm{Span}(\emptyset) = \{0\}$.
Note that $S \subseteq \mathrm{Span}(S)$, since for every $u \in X$, $u = 1 \cdot u \in \mathrm{Span}(S)$.

Example. Consider the vectors $u_1 = (1,0,0)$, $u_2 = (0,1,0)$ in $\mathbb{R}^3$. Let $S = \{u_1, u_2\} \subseteq V$.
Then vectors in $\mathrm{Span}(S)$ are precisely the vectors of the form $a_1 u_1 + a_2 u_2$ where a_1, a_2 vary over $\mathbb{R}$.
That is, $\mathrm{Span}(S)$ consists of all the vectors of the form $a_1(1,0,0) + a_2(0,1,0) = (a_1, a_2, 0)$ for some $a_1, a_2 \in \mathbb{R}$.
Thus $\mathrm{Span}(S) = \{(a_1, a_2, 0) : a_1, a_2 \in \mathbb{R}\}$ — we already know that this is a subspace of V.
This is not a coincidence!

Theorem 1.5. Let V be a vector space over F.
1) The span of any subset S of V is a subspace of V.
2) Any subspace of V that contains S must also contain $\mathrm{Span}(S)$.
(so $\mathrm{Span}(S)$ is the smallest subspace of V that contains S)

Proof. Both (1) and (2) are obvious if $S = \emptyset$, because $\mathrm{Span}(\emptyset) = \{0\}$ — we know that it is a subspace of V, and any subspace of V must contain 0.
If $S \neq \emptyset$, then S contains a vector z. As $0z = 0$, $0 \in \mathrm{Span}(S)$.
Let $x, y \in \mathrm{Span}(S)$. Then we can write
$x = a_1 u_1 + a_2 u_2 + \ldots + a_m u_m$ and $y = b_1 v_1 + \ldots + b_n v_n$ for some $a_1, \ldots, a_n, b_1, \ldots, b_n \in F$ and $u_1, \ldots, u_m, v_1, \ldots, v_n \in S$.

Then both
$x+y = a_1u_1 + \ldots + a_m u_m + b_1 v_1 + \ldots + b_n v_n$ and $c \cdot x = (ca_1)u_1 + (ca_2)u_2 + \ldots + (ca_m)u_m$
are also linear combinations of vectors of S, and so belong to Span(S).
Thus (a),(b),(c) in Theorem 1.3 hold and it follows that Span(C) is a subspace of V, showing (1).
Now let $W \leq V$ be any subspace of V that contains S.
If $w \in$ Span(S) then we can write
$w = c_1 w_1 + \ldots + c_k w_k$
for some vectors $w_1, \ldots, w_k \in S$ and some scalars $c_1, \ldots, c_k$ in F.
Since $S \subseteq W$, we have $w_1, \ldots, w_k \in W$. (W is closed under addition and multiplication)
But as W is a vector space, this implies that $w = c_1 w_1 + \ldots + c_k w_k$ is also in W.
Because w, an arbitrary vector in Span(S), belongs to W, it follows that Span$(S) \subseteq W$.
This proves (2).

Definition. A subset S of a vector space V generates (or spans) V if Span$(S) = V$.
(In this case, we also say that the vectors of S generate, or span, V.)

- Finding a small (and explicit) generating set for a vector space is an efficient way of describing V and simplifies working with it.

Example For any vector space V, Span$(V) = V$ (so V is generated by itself).

Example. The vectors $(1,0,0), (0,1,0), (0,0,1)$ generate the vector space $\mathbb{R}^3$.
Indeed, any vector $(a,b,c) \in \mathbb{R}^3$ can be written as $a\cdot(1,0,0) + b(0,1,0) + c(0,0,1)$, and so
Span$(\{(1,0,0),(0,1,0),(0,0,1)\}) = \mathbb{R}^3$.

Example. Let $V = M_{2\times 2}(\mathbb{R})$ be the vector space of all 2×2 matrices with entries from $\mathbb{R}$.
Then $M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $M_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $M_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $M_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ generate $M_{2\times 2}(\mathbb{R})$.
Indeed, for any $a,b,c,d \in \mathbb{R}$ we can write

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus Span$(\{M_1, \ldots, M_4\}) = M_{2\times 2}(\mathbb{R})$.

Example. Let $P(F)$ be the vector space of all polynomials over F.
Then the set $\{1, x, x^2, x^3, \ldots\}$ generates $P(F)$.
Indeed, Span$(\{1, x, x^2, \ldots\}) = \{a_0 + a_1 x + \ldots + a_n x^n : n \in \mathbb{N}, a_i \in F\}$ — all polynomials over F appear.
Similarly, $P_n(F)$ is generated by $\{1, x, \ldots, x^n\}$.

Linear Independence

Usually there are many subsets that generate the same space.

Example.
We saw that the vectors $(1,0,0), (0,1,0), (0,0,1)$ generate the vector space $\mathbb{R}^3$.
The set $\{(1,0,0),(0,1,0),(0,0,1),(2,3,-1)\}$ also generates $\mathbb{R}^3$, but in fact the vector $(2,3,-1)$ is redundant.

- It is natural to look for the smallest possible subset of V that generates it.
First, we explore the circumstances under which a vector can be removed from a generating set to obtain a smaller generating set.

- If $u_1, \ldots, u_n$ are any vectors in a vector space V over F, then the zero vector is always a linear combination of $u_1, \ldots, u_n$:
$0 = 0\cdot u_1 + 0\cdot u_2 + \ldots + 0\cdot u_n$.

(this is called a trivial representation of 0 as a linear combination of $u_1, \dots, u_n$).

- Sometimes it is also possible to write
$0 = a_1 u_1 + \dots + a_n u_n$
so that not all of $a_1, \dots, a_n \in F$ are 0. (a non-trivial representation of 0).

Example. In $\mathbb{R}^2$, $0 = 2\cdot(1,2) + 5(2,1) + 3(-4,-3)$ is a non-trivial representation of 0.

Definition. A subset S of a vector space V is called linearly dependent if there exist a finite number of distinct vectors $u_1, \dots, u_n$ in S and scalars $a_1, \dots, a_n \in F$, not all zero, such that $a_1 u_1 + a_2 u_2 + \dots + a_n u_n = 0$.
If S is not linearly dependent, then S is called linearly independent.
(We also say that the vectors $v_1, \dots, v_n$ are linearly dependent / independent if the set $\{v_1, \dots, v_n\}$ is linearly dependent / independent.)

Example. Let $V = \mathbb{R}^2$.

- The set $S_1 = \{(0,1), (1,0)\}$ is linearly independent.
Indeed, if $(0,0) = a_1(0,1) + a_2(1,0)$, then $\begin{cases} 0 = a_1\cdot 0 + a_2 \cdot 1 \\ 0 = a_1 \cdot 1 + a_2 \cdot 0 \end{cases}$ must hold, and so $a_1 = 0$, $a_2 = 0$. This means that any representation of the zero vector as a linear combination of vectors from S_1 is trivial.
- However, the set $S_2 = \{(0,1), (1,0), (17,18)\}$ is linearly dependent.
Indeed, $18(0,1) + 17(1,0) + (-1)(17,18) = 0$, and this is a non-trivial representation of 0.

Example. Let V be an arbitrary vector space over F.

1) Any set $S \subseteq V$ containing 0 is linearly dependent. (indeed, as $0 \in S$, then $0 = 1 \cdot 0$ is a non-trivial representation of 0.
2) The empty set $\emptyset \subseteq V$ is linearly independent (we cannot form any linear combination at all using its elements)
3) If $S = \{u\} \subseteq V$ consists of a single non-zero vector u, then S is linearly independent.
Indeed, if $\{u\}$ is linearly dependent, then $au = 0$ for some <u>non-zero</u> scalar $a \in F$. Thus
$u = \underbrace{(a^{-1}\cdot a)}_{=1} u = a^{-1}(au) = a^{-1} \cdot 0 = 0$ (by Theorem 1.2)

Example. Similarly, in $V = M_{2\times 2}(\mathbb{R})$, the set $S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ is linearly independent.
Indeed, assume that
$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = a_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + a_3 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + a_4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ — a representation of the zero vector $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$.
This means that the system of linear equations

$$\begin{cases} 0 = a_1 \cdot 1 + \cancel{a_2 \cdot 0} + \cancel{a_3 \cdot 0} + \cancel{a_4 \cdot 0} \\ 0 = \cancel{a_1 \cdot 0} + a_2 \cdot 1 + \cancel{a_3 \cdot 0} + \cancel{a_4 \cdot 0} \\ 0 = \cancel{a_1 \cdot 0} + \cancel{a_2 \cdot 0} + a_3 \cdot 1 + \cancel{a_4 \cdot 0} \\ 0 = \cancel{a_1 \cdot 0} + \cancel{a_2 \cdot 0} + \cancel{a_3 \cdot 0} + a_4 \cdot 1 \end{cases}$$

is satisfied. But this is only possible when $a_1 = a_2 = a_3 = a_4 = 0$. Thus, there are no non-trivial representation of 0 using elements from S.

Example. In $V = P_n(F)$, the set $S = \{1, x, \dots, x^n\}$ is linearly independent.
Indeed, assume that
$0 = a_0 \cdot 1 + a_1 \cdot x + \dots + a_n x^n$ is a representation of the zero vector in $P_n(F)$ (which is the zero polynomial).
This is only possible when all of a_i, $i = 0, \dots, n$ are 0, which implies that the representation is trivial.

Theorem 1.6. Let V be a vector space over F, and let $S_1 \subseteq S_2 \subseteq V$ be two subsets.

1) S_1 is linearly dependent $\Rightarrow$ S_2 is linearly dependent.
2) S_2 is linearly independent $\Rightarrow$ S_1 is linearly independent.

Proof (2) follows from (1).

To see (i), notice that if $a_1u_1 + \dots + a_nu_n$ is a linear combination of elements of S_1, then it is also a linear combination of elements of S_2.

Now we connect the notions of span and linear independence.

Theorem 1.7. Let S be a linearly independent subset of a vector space V, and let v be a vector in V that is not in S.
Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \mathrm{Span}(S)$.

Proof.
If the set $S \cup \{v\}$ is linearly dependent then we can write
$0 = a_1u_1 + \dots + a_nu_n$
for some $u_1, \dots, u_n \in S \cup \{v\}$ and some non-zero scalars $a_1, \dots, a_n \in F$.
Because S is linearly independent, it is not possible that $u_i \in S$ for all $i = 1, \dots, n$.
Thus one of the u_i, let's say u_1, equals v. Then $a_1v + a_2u_2 + \dots + a_nu_n = 0$. As $a_1 \neq 0$, we get:

$$v = \frac{1}{a_1}\left(-a_2u_2 - \dots - a_nu_n\right) = \left(-\frac{a_2}{a_1}\right)u_2 + \left(-\frac{a_3}{a_1}\right)u_3 + \dots + \left(-\frac{a_n}{a_1}\right)u_n.$$

As $\left(-\frac{a_i}{a_1}\right)$ are scalars in F for all $i = 2, \dots, n$, it follows that v is a linear combination of the vectors $u_2, \dots, u_n \in S$. Thus $v \in \mathrm{Span}(S)$.

Conversely, let $v \in \mathrm{Span}(S)$. Then we can write
$v = b_1v_1 + \dots + b_mv_m$
for some vectors $v_1, \dots, v_m \in S$ and some scalars $b_1, \dots, b_m \in F$. Hence
$0 = b_1v_1 + \dots + b_mv_m + (-1)v$.
Since v is not in S, in particular $v \neq v_i$ for $i = 1, \dots, m$.
Thus, the coefficient of v in this linear combination is non-zero, and so the set $\{v_1, \dots, v_m, v\}$ is linearly dependent.
As $\{v_1, \dots, v_m, v\} \subseteq S \cup \{v\}$, the set $S \cup \{v\}$ is linearly dependent by Theorem 1.6.

It follows that if S generates a subspace W and no proper subset of S generates W, then S is linearly indep.

Bases and dimension

Definition. A **basis** β for a vector space V is a linearly independent subset of V that generates V.
If β is a basis for V, we also say that the vectors of β form a basis for V.

Example 1. Recall that $\mathrm{Span}(\emptyset) = \{0\}$ and $\emptyset$ is linearly independent. Thus the empty set $\emptyset$ is a basis for the zero vector space.

Example 2. In $V = F^n$, let $e_1 = (1, 0, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, $\dots$, $e_n = (0, 0, \dots, 0, 1)$.
Then $\{e_1, e_2, \dots, e_n\}$ is a basis for F^n and is called the standard basis for F^n.
(In the previous exercises we have already checked this for $\mathbb{R}^2$).

Example 3. In $V = M_{m\times n}(F)$, let E^{ij} denote the matrix whose only nonzero entry is a 1 in the i^{th} row and j^{th} column.
Then the set $\{E^{ij} : 1 \le i \le m, 1 \le j \le n\}$ is a basis for $M_{m\times n}(F)$.
(Again, in the previous exercises we have already checked this for $M_{2\times 2}(\mathbb{R})$.)

$$E^{ij} = \begin{array}{cc} & \begin{array}{ccc} & j & \end{array} \\ i & \begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & 0 \end{pmatrix} \end{array}$$

Example 4. In $V = P_n(F)$, the set $\{1, x, x^2, \dots, x^n\}$ is a basis. We call this basis the standard basis for $P_n(F)$.

Example 5. In $V = P(F)$, the set $\{1, x, x^2, x^3, \dots\}$ is a basis.
This shows in particular that a basis need not be finite.
(Later we will see that in fact no basis for $P(F)$ can be finite.)

The next theorem establishes the most significant property of a basis:
- Every vector in V can be expressed in one and only one way as a linear combination of the vectors in the basis.

It is this property that makes bases the building blocks of vector spaces.

Theorem 1.8. Let V be a vector space and $\beta = \{u_1, \dots, u_n\}$ be a subset of V. Then the following two statements are equivalent.

1) β is a basis for V

2) Every vector $v \in V$ can be uniquely expressed as a linear combination of vectors in β, that is, can be expressed in the form $v = a_1u_1 + \dots + a_nu_n$ for unique scalars $a_1, \dots, a_n$.

Proof. (1) implies (2).

Let β be a basis for V. If $v \in V$ is any vector in V, then $v \in \text{Span}(\beta)$ because $\text{Span}(\beta) = V$ (by assumption). Thus v is a linear combination of the vectors in β.

Suppose that
$v = a_1u_1 + a_2u_2 + \dots + a_nu_n$ and $v = b_1u_1 + b_2u_2 + \dots + b_nu_n$
are any two such representations of v.

Substracting the 2^{nd} equation from the 1^{st} one gives:
$0 = (a_1 - b_1)u_1 + (a_2 - b_2)u_2 + \dots + (a_n - b_n)u_n$.

Since β is linearly independent, it follows that
$a_1 - b_1 = 0, \dots, a_n - b_n = 0$ (otherwise we would get a non-trivial representation of 0 with vectors in β).

Hence $a_1 = b_1, \dots, a_n = b_n$ — which means that there is a unique way to express v as a lin. comb. of the vectors in β.

(2) implies (1).

Suppose every $v \in V$ can be uniquely expressed as a lin. comb. of $u_1, \dots, u_n$.
Then $\text{Span}(\beta) = V$ (in particular), and it remains to check that β is linearly independent.
Assume $0 = a_1u_1 + \dots + a_nu_n$ for some $a_1, \dots, a_n$ in F.
But also $0 = 0 \cdot u_1 + \dots + 0 \cdot u_n$.
These are two ways to express $0 \in V$, so by the uniqueness assumption they must coincide.
That is, we must have $a_1 = 0$, $a_2 = 0$, $\dots$, $a_n = 0$ — which means that β is lin. indep.

Basis and dimension.

Recall that $\beta \subseteq V$ is a basis for V if $\mathrm{Span}(\beta) = V$ and β is a lin. indep. set.
We have shown (Theorem 1.8) that if $\beta = \{u_1, \dots, u_n\}$ is a basis for V, then every vector $v \in V$ can be expressed as
$v = a_1 u_1 + \dots + a_n u_n$
for a unique choice of the scalars $a_1, \dots, a_n \in F$.

- But how does one find a basis for V?

Theorem 1.9. If V is a vector space generated by a finite set S, then some subset of S is a basis for V. In particular, V has a finite basis.

Proof.
If $S = \emptyset$ or $S = \{0\}$, then $V = \mathrm{Span}(S) = \{0\}$ and $\emptyset$ is a subset of S that is a basis for V.
Otherwise, S contains a vector $u_1 \neq 0$.
By the previous example, the set $\{u_1\}$ is linearly independent.
If there is u_2 in S s.t. $\{u_1, u_2\}$ is still linearly indep., add it to $\{u_1\}$ to get $\{u_1, u_2\}$.
If there is $u_3 \in S$ $\{u_1, u_2, u_3\}$ is lin. indep., add it to obtain the set $\{u_1, u_2, u_3\}$, etc...
Since S is finite, this process must stop on some step n, and we obtain a set
$\beta = \{u_1, \dots, u_n\} \subseteq S$ s.t. β is linearly indep., but $\beta \cup \{x\}$ is lin. dependent for any $x \in S \setminus \beta$.

Claim. β is a basis for V.
- β is lin. indep. — by construction.
- Remains to show: $\mathrm{Span}(\beta) = V$.

By Theorem 1.5, need to show that $S \subseteq \mathrm{Span}(\beta)$ — as $\mathrm{Span}(\beta)$ is a subspace of V containing S, it must also contain $\mathrm{Span}(S) = V$.

Let $v \in S$ be arbitrary.
If $v \in \beta$, then $v \in \mathrm{Span}(\beta)$.
Otherwise, if $v \notin \beta$, then by construction $\beta \cup \{v\}$ is lin. dep. — so $v \in \mathrm{Span}(\beta)$ by Theorem 1.7.
Thus $S \subseteq \mathrm{Span}(\beta)$.

- Existence of a basis in V can be proved without assuming that S is finite as well, but the proof is more involved.
- Thus, any finite generating set for V can be reduced to a basis for V, by removing some vectors.

Example. The set $S = \{(2,-3,5), (1,0,-2), (7,2,0), (0,1,0)\} \subseteq \mathbb{R}^3$ generates $\mathbb{R}^3$ (check it!)
We reduce it to a basis of $\mathbb{R}^3$, as in the proof of Theorem 1.9.
$S_0 = \{(2,-3,5)\}$ — lin. indep.
$S_1 = \{(2,-3,5), (1,0,-2)\}$ ← still lin. indep. (check it!)
$S_2 = \{(2,-3,5), (1,0,-2), (7,2,0)\}$ ✓
But $(0,1,0) \in \mathrm{Span}(S_2)$: $-\frac{7}{30}(2,-3,5) - \frac{35}{60}(1,0,-2) + \frac{3}{20}(7,2,0) = (0,1,0)$.
Hence $\beta = S_2$ is a basis for $\mathbb{R}^3$.

Now, the key technical result of this section.

Theorem 1.10 (Replacement). Let V be a v.s. generated by a set $G \subseteq V$ with $|G| = n$ (the size of G), and let $L \subseteq V$ be a lin. indep. subset of V with $|L| = m$.
Then $m \leq n$, and there exists $H \subseteq G$ with $|H| = n - m$ such that $L \cup H$ generates V.

Proof. We prove it by induction on m.
For $m = 0$, $L = \emptyset$, and so we can take $H = G$.
Now suppose the result is true for $m \geq 0$, and we prove it for $m+1$.
Let $L = \{v_1, \dots, v_{m+1}\} \subseteq V$ be lin. indep., $|L| = m+1$.
By Theorem 1.6, $\{v_1, \dots, v_m\}$ is also lin. indep. Applying the induction hypothesis, $m \leq n$ and there is a subset $\{u_1, \dots, u_{n-m}\} \subseteq G$ s.t. $\{v_1, \dots, v_m\} \cup \{u_1, \dots, u_{n-m}\}$ generates V.

So, there exist $a_1, \dots, a_m, b_1, \dots, b_{n-m}$ such that

$$a_1 v_1 + \dots + a_m v_m + b_1 u_1 + \dots + b_{n-m} u_{n-m} = v_{m+1} \qquad (*)$$

Note. Since $\{v_1, \dots, v_m, v_{m+1}\}$ is lin. indep., we must have $n > m$ (that is, $n \geq m+1$) and some $b_i \neq 0$, say $b_1 \neq 0$.
(otherwise v_{m+1} is a lin. combination of $v_1, \dots, v_m$).

Solving $(*)$ for u_1 gives:

$$u_1 = \left(-\frac{a_1}{b_1}\right) v_1 + \dots + \left(-\frac{a_m}{b_1}\right) v_m + \left(\frac{1}{b_1}\right) v_{m+1} + \left(-\frac{b_2}{b_1}\right) u_2 + \dots + \left(-\frac{b_{n-m}}{b_1}\right) u_{n-m}. \qquad (**)$$

Let $H = \{u_2, \dots, u_{n-m}\}$, so $|H| = n-(m+1)$.

Then $u_1 \in \mathrm{Span}(L \cup H)$ by $(**)$, and so $\{v_1, \dots, v_m, u_1, \dots, u_{n-m}\} \subseteq \mathrm{Span}(L \cup H)$.

As $\{v_1, \dots, v_m, u_1, \dots, u_{n-m}\}$ generates V, $\mathrm{Span}(L \cup H) = V$ (by Theorem 1.5)

Thus, the theorem is true for $m+1$.

This theorem has very important consequences.

Corollary 1. Let V be a v.s. having a finite basis. Then every basis for V contains the same number of vectors.

Proof. Suppose $\beta \subseteq V$ with $|\beta| = n$ is a basis for V, and let $\gamma \subseteq V$ be any other basis for V.

Suppose that $|\gamma| > n$, and let $S \subseteq \gamma$ have $n+1$ elements.

Since S is lin. indep. and β generates V, by replacement $n+1 \leq n$ — a contradiction.

Thus $|\gamma| = m \leq n$.

Reversing the roles of β and γ, by the same argument we get $n \leq m$. Hence $n = m$.

This fact makes possible the following important definition.

Definition. A v.s. V is finite-dimensional if it has a finite basis.
The (unique) number of vectors in a basis for V is called the dimension of V, denoted $\dim(V)$.
If there is no finite basis, then V is infinite-dimensional.

Example. In view of the previous discussion, we have:

1) $\dim(\{0\}) = 0$. ($\emptyset$ is the basis).
2) $\dim(F^n) = n$. ($\{(1,0,\dots,0), (0,1,0,\dots,0), \dots, (0,\dots,0,1)\}$ is a basis of size n).
3) $\dim(M_{m \times n}) = mn$. ($\{E^{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis of size mn; recall that $E^{ij} = \begin{array}{c c} & j \\ i & \begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & 0 \end{pmatrix} \end{array}$).
4) $\dim(P_n(F)) = n+1$. ($\{1, x, x^2, \dots, x^n\}$ is a basis of size $n+1$).

Example. On the other hand, some of the familiar examples are infinite-dimensional.
By the replacement theorem, if V is finite-dimensional, then no lin. indep. set can contain more than $\dim(V)$ elements. Thus:

$P(F)$ is infinite-dimensional (as $\{1, x, x^2, x^3, \dots\}$ is an infinite lin. indep. set).

Corollary 2. Let V be a v.s. of dimension n.

a) Any lin. indep. subset of V with n elements is a basis.
b) Every lin. indep. subset of V can be extended to a basis for V.

Proof. Let β be a basis for V, $|\beta| = n$.

a) Let $L \subseteq V$ be lin. indep. with $|L| = n$. By the Replacement theorem, $\exists H \subseteq \beta$ with $|H| = n - n = 0$ elements such that $L \cup H$ generates V. Thus $H = \emptyset$, and so L gener V — so L is a basis.

b) If $L \subseteq V$ is lin. indep. with $|L| = m$, by the Replacement theor $\exists H \subseteq \beta$ with $|H| = n - m$ such that $L \cup H$ generates V. Now $|L \cup H| \leq m + (n-m) = n$.
By Theorem 1.9 $L \cup H$ contains some subset γ which is a basis for V, and $|\gamma| = n$ by Corollary 1.
But then $\gamma = L \cup H$.

Theorem 1.11.
Let W be a subspace of a v.s. V with $\dim(V) < \infty$.
Then $\dim(W) \leq \dim(V)$.
Moreover, if $\dim(W) = \dim(V)$, then $V = W$.

Proof.

Let $\dim(V) = n$.

If $W = \{0\}$ then $\dim(W) = 0 \le n$ (by the previous example).

Otherwise $\exists\, x_1 \in W$, $x_1 \neq 0$. So $\{x_1\}$ is a lin. indep. set.

Continue choosing $x_1, \dots, x_k \in W$ s.t. $\{x_1, \dots, x_k\}$ is lin. indep.

Since no lin. indep. subset of V can contain more than n vectors (Cor 1 + Cor 2), this process must stop at a stage where:

$k \le n$, $\{x_1, \dots, x_k\}$ is lin. indep., but $\{x_1, \dots, x_k\} \cup \{v\}$ is lin. dep. for any $v \in W$.

By Theorem 1.7, this implies $\mathrm{Span}(\{x_1, \dots, x_k\}) = W$, hence $\{x_1, \dots, x_k\}$ is a basis for W.

So $\dim(W) = k \le n$.

(and by Corollary 2(a), if $k = n$ then $\{x_1, \dots, x_k\}$ is a basis for V, hence $W = V$.)

Corollary. If W is a subspace of a v.s. V with $\dim(V) < \infty$, then any basis for W can be extended to a basis for V.

Proof. If $S \subseteq W$ is a basis for W, it is a lin. indep. subset of V, so can be extended to a basis for V.

Example.

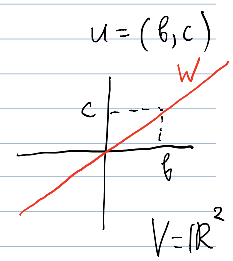
1) Let's describe all subspaces of $V = \mathbb{R}^2$.

We know $\dim(\mathbb{R}^2) = 0$ (as $\{(1,0), (0,1)\}$ is a basis).

By Theorem 1.11, for every subspace $W \le \mathbb{R}^2$ we must have $\dim(W) = 0, 1$ or 2.

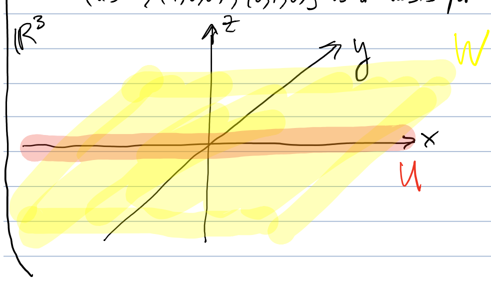
If $\dim(W) = 0$ then $W = \{0\}$ and if $\dim(W) = 2$ then $W = \mathbb{R}^2$.

And if $\dim(W) = 1$, then $W = \{a \cdot u : a \in F\}$ for some non-zero vector $u \in \mathbb{R}^2$.



2) If $V = \mathbb{R}^3$, then $\dim(V) = 3$, and for $W = \{(a, b, 0) : a, b \in \mathbb{R}\}$ we have $\dim(W) = 2$.

(as $\{(1,0,0), (0,1,0)\}$ is a basis for W) and $\dim(U) = 1$ for $U = \{(a, 0, 0) : a \in \mathbb{R}\}$.



W is the xy-plane

U is the x-axis

We can list all subspaces of $\mathbb{R}^3$:

$\dim(W) = 0$ — W is the origin point,

$\dim(W) = 1$ — W is a line through the origin,

$\dim(W) = 2$ — W is a plane through the origin,

$\dim(W) = 3$ — $W = \mathbb{R}^3$.

Linear transformations

Definition. Let V and W be v.s. (over F).

A function $T: V \to W$ is a *linear transformation from V to W* if, for all $x, y \in V$ and $c \in F$:

(a) $T(x + y) = T(x) + T(y)$

(b) $T(cx) = c\,T(x)$

addition and scalar mult. in V

addition and scalar mult. in W.

Basic properties of linear transformations

Let $T: V \to W$ be a lin. transformation. Then:

1) $T(0) = 0$

2) $T(cx + y) = c\,T(x) + T(y)$ for all $x, y \in V$, $c \in F$. (This holds if and only if T is linear).

3) $T(x - y) = T(x) - T(y)$

4) $T\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i T(x_i)$ for all $x_i \in V$, $a_i \in F$.

Proof. Exercise.

Example. Some examples of lin. transformations $T: \mathbb{R}^2 \to \mathbb{R}^2$.

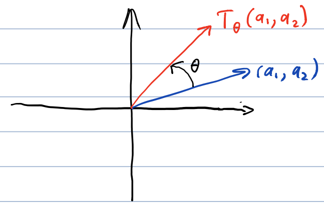
1) $T(a_1, a_2) = (5a_1, 3a_2)$.

2) For any $\theta \in \mathbb{R}$, define:

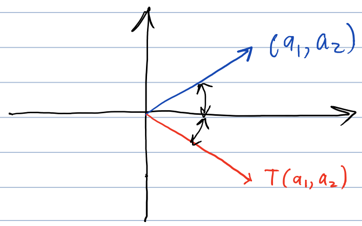
$T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$

$T_\theta(a_1, a_2) = (a_1 \cos\theta - a_2 \sin\theta,\ a_1 \sin\theta + a_2 \cos\theta)$. — check that T is linear!

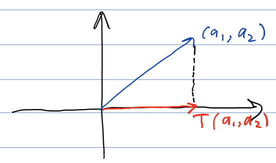
– the rotation (counter-clockwise) by the angle θ.



3) $T(a_1, a_2) = (a_1, -a_2)$ – the reflection about the x-axis.



4) $T(a_1, a_2) = (a_1, 0)$ – the projection on the x-axis.



Example. We define $T: M_{m\times n}(F) \to M_{n\times m}(F)$ by $T(A) = A^t$, where A^t is the transpose of A.
Then T is a lin. transformation.

Example. Define $T: P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ by $T(f(x)) = f'(x)$, where $f'(x)$ denotes the derivative of $f(x)$.
To show that T is linear, let $g(x), h(x) \in P_n(\mathbb{R})$ and $a \in \mathbb{R}$ be arbitrary. Then:
$T(ag(x) + h(x)) = (ag(x) + h(x))' = ag'(x) + h'(x) = a \cdot T(g(x)) + T(h(x))$.

Example. Let $V = C(\mathbb{R})$, the vector space of continuous real-valued functions on $\mathbb{R}$.
Let $a, b \in \mathbb{R}$, $a < b$ be fixed. We define $T: V \to \mathbb{R}$ (the v.s. $\mathbb{R}^1$) by:
$T(f) = \int_a^b f(t)\,dt$
for all functions $f \in V$.
Then T is linear (because $\int_a^b (ag(t) + h(t))\,dt = a\int_a^b g(t)\,dt + \int_a^b h(t)\,dt = a \cdot T(g) + T(h)$.

Null space and range.

Definition. Let V and W be v.s., and $T: V \to W$ be linear.
1) Let $N(T) = \{x \in V : T(x) = 0\}$ – the null space (or kernel) of T.
2) Let $R(T) = \{T(x) : x \in V\}$ – the range (or image) of T.

Example. Let V and W be v.s.
1) We define $I: V \to V$ by $I(x) = x$ for all $x \in V$ — the identity transformation.
Then I is linear, $N(I) = \{0\}$ and $R(I) = V$.
2) We define $T_0 : V \to W$ by $T_0(x) = 0$ for all $x \in V$ — the zero transformation.
Then T_0 is linear, $N(T_0) = V$ and $R(T_0) = \{0\}$.

Theorem 2.1. Let V, W be v.s and $T: V \to W$ linear.
Then $N(T)$ and $R(T)$ are subspaces of V and W, respectively.

Proof.

1) $N(T)$ is a subspace of V.
   (a) $0 \in N(T)$ — as $T(0) = 0$.
   b), c) Let $x, y \in N(T)$ and $c \in F$.
   Then $T(x+y) = T(x) + T(y) = 0 + 0 = 0$ and $T(cx) = c \cdot T(x) = c \cdot 0 = 0$.
   Hence $x+y \in N(T)$ and $cx \in N(T)$.
   So $N(T)$ is a subspace of V.
2) $R(T)$ is a subspace of W.
   Analogous (do it!).

Theorem 2.2 Let V, W be v.s. and $T: V \to W$ linear.
If $\beta = \{v_1, \dots, v_n\}$ is a basis for V, then
$R(T) = \text{Span}(T(\beta)) = \text{Span}(\{T(v_1), \dots, T(v_n)\})$.

Proof. Clearly $T(v_i) \in R(T)$ for each i.
As $R(T)$ is a subspace of W, $\text{Span}(\{T(v_1), \dots, T(v_n)\}) = \text{Span}(T(\beta)) \subseteq R(T)$ (by Theorem 1.5)
Suppose $w \in R(T)$, then $w = T(v)$ for some $v \in V$.
As β is a basis for V, we have
$v = \sum_{i=1}^{n} a_i v_i$ for some $a_i \in F$.
And since T is linear,
$w = T(v) = \sum_{i=1}^{n} a_i T(v_i) \in \text{Span}(T(\beta))$.
Hence $R(T) \subseteq \text{Span}(T(\beta))$.

Definition. Let V, W be v.s. and $T: V \to W$ linear.
If $N(T), R(T)$ are finite-dimensional, then we define
nullity$(T) = \dim(N(T))$,
rank$(T) = \dim(R(T))$.

- Intuitively, if $N(T)$ is "large" (that is, T sends many vectors from V to 0), then $R(T)$ should be "small" (not so many vectors in W can be obtained by T from the vectors in V). And vice versa.

Theorem 2.3 (Dimension Theorem). Let V, W be v.s. and $T: V \to W$ linear. If $\dim(V) < \infty$ then
nullity(T) + rank$(T) = \dim(V)$.

Proof.
Suppose that $\dim(V) = n$, $\dim(N(T)) = k$, and $\{v_1, \dots, v_k\}$ is a basis for $N(T)$.
By the Corollary to Theorem 1.11:
Can extend $\{v_1, \dots, v_k\}$ to a basis $\beta = \{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ for V.

Claim. $S = \{T(v_{k+1}), \dots, T(v_n)\}$ is a basis for $R(T)$.

- S generates $R(T)$.
  As $T(v_i) = 0$ for $1 \le i \le k$, by Theorem 2.2:
  $R(T) = \text{Span}(\{T(v_1), \dots, T(v_n)\}) = \text{Span}(\{T(v_{k+1}), \dots, T(v_n)\}) = \text{Span}(S)$.
- S is lin. indep.:
  Suppose $\sum_{i=k+1}^{n} b_i T(v_i) = 0$ for $b_{k+1}, \dots, b_n \in F$.
  As T is linear, $T\left(\sum_{i=k+1}^{n} b_i v_i\right) = 0$.
  So $\sum_{i=k+1}^{n} b_i v_i \in N(T)$.
  Hence $\exists\, c_1, \dots, c_k \in F$ such that $\sum_{i=k+1}^{n} b_i v_i = \sum_{i=1}^{k} c_i v_i$, (as $\{v_1, \dots, v_k\}$ is a basis for $N(T)$). or $\sum_{i=1}^{k} (-c_i) v_i + \sum_{i=k+1}^{n} b_i v_i = 0$.
  Since β is a basis for V, we have $b_i = 0$ for all i.
  Hence S is lin. indep.

So $\dim(V) = n$, $\dim(N(T)) = k$ and $\dim(R(T)) = n - k$.

Properties of lin. transformations (contd.)

Example.

1) Let $T: F^n \to F^{n-1}$ be defined by $T((a_1,\dots,a_n)) = (a_1,\dots,a_{n-1})$. — so T "forgets" the n-th component.
Then T is linear, $N(T) = \{(\underbrace{0,\dots,0}, a_n) : a_n \in F\}$ and $R(T) = F^{n-1}$.
And $\dim(F^n) = n$, $\dim(N(T)) = 1^{n-1}$ and $\dim(R(T)) = \dim(F^{n-1}) = n-1$.

2) Let $T: P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ be the differentiation transformation, that is $T(p(x)) = p'(x)$ for any polynomial $p(x)$.
Then $T(p(x)) = 0 \iff p'(x) = 0 \iff p(x)$ constant. So $N(T) = \{$ constant polynomials in $P(\mathbb{R})\}$.
Recall that $\{1, x, x^2, \dots, x^{n-1}\}$ is a basis for $P_{n-1}(\mathbb{R})$. Since $1 = T(x)$, $x = \frac{1}{2}T(x^2), \dots, x^{n-1} = \frac{1}{n}T(x^n)$, it follows that $V = \operatorname{Span}(\{T(x), \dots, T(x^n)\}) = R(T)$.
Thus $\dim(P_n(\mathbb{R})) = n+1$, $\dim(R(T)) = n$ and $\dim(N(T)) = 1$.

Definition. Let $T: V \to W$ be a lin. transf.
T is **injective** if $T(v) = T(u)$ implies $v = u$, for all $u, v \in V$.
T is **surjective** if for every $w \in W$ there is some $v \in V$ such that $T(v) = w$.
T is **bijective** if it is both injective and surjective.

Theorem 2.4 Let $T: V \to W$ be linear. Then T is injective if and only if $N(T) = \{0\}$.

Proof.
"$\Rightarrow$" Suppose T is injective, and let $x \in N(T)$.
Then $T(x) = 0 = T(0) \Rightarrow x = 0$. Hence $N(T) = \{0\}$.
"$\Leftarrow$". Assume $N(T) = \{0\}$ and suppose $T(x) = T(y)$.
Then $0 = T(x) - T(y) = T(x-y)$, as T is lin.
So $x - y \in N(T) = \{0\}$. Hence $x - y = 0$, or $x = y$.

Theorem 2.5. Let $T: V \to W$ be lin, and $\dim(V) = \dim(W) < \infty$. Then the following are equivalent:
a) T is injective.
b) T is surjective.
c) T is bijective.
d) $\dim(R(T)) = \dim(V)$.

Proof.
By the dimension theorem, $\dim(N(T)) + \dim(R(T)) = \dim(V)$.
We have:
T is injective $\overset{\text{(Theorem 2.4)}}{\iff} N(T) = \{0\} \iff \dim(N(T)) = 0 \iff \dim(R(T)) = \dim(V) \iff$
$\iff \dim(R(T)) = \dim(W) \underset{\text{(Thm 1.11)}}{\iff} R(T) = W \iff T$ is surjective.

Example.
1) Define $T: F^2 \to F^2$ by $T(a_1, a_2) = (a_1 + a_2, a_1)$.
Then $N(T) = \{0\}$, so T is injective. By Theorem 2.5, T is also surjective.
2) Define $T: P_n(\mathbb{R}) \to \mathbb{R}^{n+1}$ by $T(a_0 + a_1 x + \dots + a_n x^n) = (a_0, a_1, \dots, a_n)$.
Then T is linear and injective, hence T is bijective (as $\dim(P_n(\mathbb{R})) = \dim(\mathbb{R}^{n+1})$!).

Next we show that every lin. transf. is completely determined by its action on a basis!

Theorem 2.6.
Let V, W be v.s. over a field F, and let $\{v_1, \dots, v_n\}$ be a basis for V.
For any $w_1, \dots, w_n \in W$ there exists **exactly one** lin. transformation $T: V \to W$ s.t. $T(v_i) = w_i$ for $i = 1, \dots, n$.

Proof.

Let $x \in V$. Then $x = \sum_{i=1}^{n} a_i v_i$ for some unique scalars $a_1, \dots, a_n \in F$. (because $\{v_1, \dots, v_n\}$ is a basis!)
We define a map $T: V \to W$ by
$T(x) = \sum_{i=1}^{n} a_i w_i$.

a) T is linear.
Suppose $u, v \in V$ and $d \in F$. We can write
$u = \sum_{i=1}^{n} b_i v_i$ and $v = \sum_{i=1}^{n} c_i v_i$ for some scalars $b_1, \dots, b_n, c_1, \dots, c_n \in F$.
Then
$du + v = \sum_{i=1}^{n} (d b_i + c_i) v_i$.
So
$T(du+v) = \sum_{i=1}^{n} (d b_i + c_i) w_i = d \sum_{i=1}^{n} b_i w_i + \sum_{i=1}^{n} c_i w_i = d\, T(u) + T(v)$.

b) $T(v_i) = w_i$ for $i = 1, \dots, n$ — clear from the definition of T.

c) T is unique.
Suppose that $U: V \to W$ is linear, and that it also satisfies $U(v_i) = w_i$ for $i = 1, \dots, n$.
Then, for $x \in V$ with $x = \sum_{i=1}^{n} a_i v_i$ we have (as U is linear):
$U(x) = \sum_{i=1}^{n} a_i U(v_i) = \sum_{i=1}^{n} a_i w_i = T(x)$.
Hence $U = T$.

Corollary. Let V, W be v.s.; V has a finite basis $\{v_1, \dots, v_n\}$.
If $U, T : V \to W$ are linear and $U(v_i) = T(v_i)$ for $i = 1, \dots, n$ then $U = T$.

Example. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the lin. transformation defined by
$T(a_1, a_2) = (2a_2 - a_1, 3a_1)$
Suppose that $U: \mathbb{R}^2 \to \mathbb{R}^2$ is any lin. transf.
If we know that $U(1,2) = (3,3)$ and $U(1,1) = (1,3)$, then $U = T$.
This follows from the corollary, because $\{(1,2), (1,1)\}$ is a basis for $\mathbb{R}^2$.

The matrix representation of a lin. transformation.

Definition Let V be a fin. dim. v.s. An ordered basis for V is a basis for V endowed with a specific order.

Example.
- In F^3, $\beta = \{e_1, e_2, e_3\}$ is an ordered basis. (recall $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, $e_3 = (0,0,1)$).
  Also, $\gamma = \{e_2, e_1, e_3\}$ is an ordered basis.
  So β and γ is the same set, but $\beta \neq \gamma$ as ordered bases. The choice of the order matters!
- For the v.s. F^n, we call $\{e_1, e_2, \dots, e_n\}$ the standard ordered basis for F^n.
- Similarly, for the v.s. $P_n(F)$ we call $\{1, x, \dots, x^n\}$ the standard ordered basis for $P_n(F)$.

Definition. Let $\beta = \{u_1, \dots, u_n\}$ be an ordered basis for a fin. dim. v.s. V.
For $x \in V$, let $a_1, \dots, a_n \in F$ be the unique scalars such that
$x = \sum_{i=1}^{n} a_i u_i$. (by Theorem 1.8.)

We define the coordinate vector of x relative to β, denoted $[x]_\beta$, by

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \quad (\text{so } [x]_\beta \text{ is a vector in } F^n).$$

so each vector can be described by its coordinates with respect to a fixed basis.

- Notice that $[u_i]_\beta = e_i$.
- The correspondence $x \to [x]_\beta$ is a lin. transformation from V to F^n. (Exercise).

Example. Let $V = P_2(\mathbb{R})$, and let $\beta = \{1, x, x^2\}$ be the standard ordered basis for V.
Consider $f(x) = 4 + 6x - 7x^2 \in V$, then

$$[f]_\beta = \begin{pmatrix} 4 \\ 6 \\ -7 \end{pmatrix}.$$

Definition. Suppose V, W are fin. dim. v.s., with ordered bases $\beta = \{v_1, \dots, v_n\}$ and $\gamma = \{w_1, \dots, w_m\}$, respectively. Let $T : V \to W$ be linear.
Then for each j, $1 \le j \le n$, there exist unique scalars $a_{ij} \in F$, $1 \le i \le m$ such that
$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i$ for $1 \le j \le n$.

We call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$ the matrix representation of T in the ordered bases β and γ, and write $A = [T]_\beta^\gamma$.
If $V = W$ and $\beta = \gamma$, then we write $A = [T]_\beta$.

Notice. • the j^{th} column of A is $[T(v_j)]_\gamma$.
- If $U : V \to W$ is a lin. transf. s.t. $[U]_\beta^\gamma = [T]_\beta^\gamma$ then $U = T$ (by the corollary to Theorem 2.6)
- So $[T]_\beta^\gamma$ gives an explicit way to describe T which is very useful in computations.

Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the lin. transf. defined by
$T(a_1, a_2) = (a_1 + 3a_2,\ 0,\ 2a_1 - 4a_2)$.
Let $\beta = \{e_1, e_2\}$, $\gamma = \{e_1, e_2, e_3\}$ – the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Now:
$T(1,0) = (1,0,2) = 1e_1 + 0e_2 + 2e_3$
$T(0,1) = (3,0,-4) = 3e_1 + 0e_2 - 4e_3$.
Hence
$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}.$$
But if we take $\gamma' = \{e_3, e_2, e_1\}$, then $[T]_\beta^{\gamma'} = \begin{pmatrix} 2 & -4 \\ 0 & 0 \\ 1 & 3 \end{pmatrix}$.

Definition Let $T, U : V \to W$ be functions, where V, W are v.s. over F, and let $a \in F$. We define:
$T + U : V \to W$ by $(T+U)(x) = T(x) + U(x)$ for all $x \in V$.
$aT : V \to W$ by $(aT)(x) = aT(x)$ for all $x \in V$.
So $T+U$ and aT are both functions from V to W.

These operations preserve linearity.

Theorem 2.7 Let V, W be v.s. over F, let $T, U : V \to W$ be linear.
a) For all $a \in F$, $aT + U$ is linear.
b) With this operations of addition and scalar multiplication, the set of all linear transformations from V to W is a v.s. over F.

Proof.
a) Let $x, y \in V$ and $c \in F$. Then
$\underset{\text{function}}{(aT+U)}\ \underset{\text{vector in } V}{(cx+y)} = (aT)(cx+y) + U(cx+y) = a\big(T(cx+y)\big) + (cU)(x) + U(y) = a\big(cT(x) + T(y)\big) + cU(x) + U(y) =$
$= acT(x) + cU(x) + aT(y) + U(y) = c(aT+U)(x) + (aT+U)(y)$.
Hence the map $aT + U$ is linear.
b) Note that the zero transformation T_0 (recall $T_0(x) = 0$ for all $x \in V$) plays the role of the zero vector, and its easy to verify that all of the axioms (VS1)–(VS8) of a vector space are satisfied.

Definition For V, W v.s. over F, we denote $\mathcal{L}(V, W) = \{T : T \text{ is a lin. transf. from } V \text{ to } W\}$ – a v.s. over F.
In case $V = W$, we write $\mathcal{L}(V)$ instead of $\mathcal{L}(V, W)$.

Algebraic description of the operations in $\mathcal{L}(V,W)$.

Last time we saw:
- every lin. transformation can be represented by a matrix,
- linear transformations from V to W form a vector space $\mathcal{L}(V,W)$, under pointwise addition and scalar mult.

These operations on $\mathcal{L}(V,W)$ correspond to matrix addition and scalar mult. on the representations.

Theorem 2.8

Let V, W be fin. dim. v.s. with ordered bases β and γ, respectively.

Let $T, U : V \to W$ be linear. Then:

a) $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$. operations on matrices!

b) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ for all scalars $a \in F$.

Proof.

a) Let $\beta = \{v_1, \dots, v_n\}$ and $\gamma = \{w_1, \dots, w_m\}$.

There exist unique scalars a_{ij} and b_{ij} $(1 \le i \le m,\ 1 \le j \le n)$ s.t.:

$T(v_j) = \sum_{i=1}^m a_{ij} w_i$ and $U(v_j) = \sum_{i=1}^m b_{ij} w_i$ for $1 \le j \le n$.

Hence

$$(T+U)(v_j) = \sum_{i=1}^m (a_{ij} + b_{ij}) w_i.$$

Thus, for the matrix $[T+U]_\beta^\gamma$ we have

$$([T+U]_\beta^\gamma)_{ij} = a_{ij} + b_{ij} = ([T]_\beta^\gamma + [U]_\beta^\gamma)_{ij}.$$

b) Similar (Exercise.)

Example. Let $T, U : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$,

$U(a_1, a_2) = (a_1 - a_2, 2a_1, 3a_1 + 2a_2)$.

Let β, γ be the standard ordered bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, resp. Then

$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}$ – calculated in the previous example $\quad [U]_\beta^\gamma = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}$

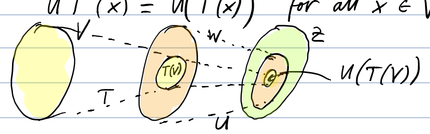
Applying definition, we have

$(T+U)(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2) + (a_1 - a_2, 2a_1, 3a_1 + 2a_2) = (2a_1 + 2a_2, 2a_1, 5a_1 - 2a_2)$. So

$[T+U]_\beta^\gamma = \begin{pmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{pmatrix} = [T]_\beta^\gamma + [U]_\beta^\gamma$ – as Theorem 2.8 predicts.

Composition of lin. transf.'s and matrix multiplication.

Definition. Let $T : V \to W$ and $U : W \to Z$ be two lin. transf.'s of v.s.'s.
Their composition, denoted by UT, is a function from V to Z defined by
$UT(x) = U(T(x))$ for all $x \in V$.



Theorem 2.9. Let V, W, Z be v.s. over F.
Let $T : V \to W$ and $U : W \to Z$ be linear.
Then $UT : V \to Z$ is linear.

Proof.

Let $x, y \in V$ and $a \in F$. Then

$$UT(ax+y) = U(T(ax+y)) \overset{(T \text{ is lin.})}{=} U(aT(x) + T(y)) \overset{(U \text{ is lin.})}{=} aU(T(x)) + U(T(y)) = a(UT)(x) + UT(y).$$

- See Problem Set 4 for more basic properties of the composition.

- Assume that V, W, Z are v.s. over F, and let
$\alpha = \{v_1, \dots, v_n\}$, $\beta = \{w_1, \dots, w_m\}$, $\gamma = \{z_1, \dots, z_p\}$ be ordered bases for V, W and Z, respectively.
Let $T: V \to W$ and $U: W \to Z$ be linear.
Let $A = [U]_\beta^\gamma$ and $B = [T]_\alpha^\beta$ be their matrix representations.
We have $UT: V \to Z$ — their composition.
- Let's calculate its matrix representation $[UT]_\alpha^\gamma$.
For $1 \le j \le n$, we have
$$(UT)(v_j) = U(T(v_j)) = U\left(\sum_{k=1}^m B_{kj}\, w_k\right) = \sum_{k=1}^m B_{kj} U(w_k) = \sum_{k=1}^m B_{kj}\left(\sum_{i=1}^p A_{ik} z_i\right) = \sum_{i=1}^p \left(\sum_{k=1}^m A_{ik} B_{kj}\right) z_i = \sum_{i=1}^p C_{ij} z_i$$
where
$$C_{ij} = \sum_{k=1}^m A_{ik} B_{kj}.$$

Hence $[UT]_\alpha^\gamma = C = (C_{ij})_{\substack{1 \le i \le p \\ 1 \le j \le n}}$.
This computation motivates the definition of matrix multiplication.

Definition. Let A be an $m \times n$ matrix, and B an $n \times p$ matrix. We define the product of A and B, denoted AB, to be the $m \times p$ matrix such that
$$(AB_{ij}) = \sum_{k=1}^n A_{ik} B_{kj} \quad \text{for } 1 \le i \le m,\ 1 \le j \le p.$$

Example.

1) $$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 4 & -1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1\cdot 4 + 2\cdot 2 + 1\cdot 5 \\ 0\cdot 4 + 4\cdot 2 + (-1)\cdot 5 \end{pmatrix} = \begin{pmatrix} 13 \\ 3 \end{pmatrix}.$$

2) Matrix multiplication is not commutative.
$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \text{ but } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$ so it is possible that $AB \ne BA$.

3) Recall the definition of the transpose of a matrix from Problem Set 2:
If $A \in M_{m\times n}(F)$, then its transpose $A^t \in M_{n\times m}(F)$ is given by $(A^t)_{ij} = A_{ji}$ for all $1 \le i \le m$, $1 \le j \le n$.
We show that
$$(AB)^t = B^t A^t$$
Indeed, we have
$$(AB)^t_{ij} = (AB)_{ji} = \sum_{k=1}^n A_{jk} B_{ki} = \sum_{k=1}^n B_{ki} A_{jk} = \sum_{k=1}^n (B^t)_{ik} (A^t)_{kj} = (B^t A^t)_{ij}.$$

Returning to our previous calculation, we can now state it in a compact form using matrix multiplication.

Theorem 2.11.
Let V, W and Z be fin. dim. v.s. with ordered bases α, β and γ, respectively.
Let $T: V \to W$ and $U: W \to Z$ be lin. transformations. Then
$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta.$$

Corollary. Let V be a fin. dim. v.s. with an ordered basis β.
Let $T, U \in \mathcal{L}(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.

Example. Let $U: P_3(\mathbb{R}) \to P_2(\mathbb{R})$ and $T: P_2(\mathbb{R}) \to P_3(\mathbb{R})$ be the lin. transf. defined by
$U(f(x)) = f'(x)$ and $T(f(x)) = \int_0^x f(t)\,dt$.
Let $\alpha = \{1, x, x^2, x^3\}$ and $\beta = \{1, x, x^2\}$ be the standard ordered bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively.
We have:

$U(1) = 0 = 0\cdot 1 + 0\cdot x + 0\cdot x^2$

$U(x) = 1 = 1\cdot 1 + 0\cdot x + 0\cdot x^2$

$U(x^2) = 2x = 0\cdot 1 + 2\cdot x + 0\cdot x^2$

$U(x^3) = 3x^2 = 0\cdot 1 + 0\cdot x + 3\cdot x^2.$

Hence $$[U]_\alpha^\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Similarly, for T we have:

$$T(1) = x = 0\cdot 1 + 1\cdot x + 0\cdot x^2 + 0\cdot x^3$$
$$T(x) = \tfrac{1}{2}x^2 = 0\cdot 1 + 0\cdot x + \tfrac{1}{2}x^2 + 0\cdot x^3$$
$$T(x_2) = \tfrac{1}{3}x^3 = 0\cdot 1 + 0\cdot x + 0\cdot x^2 + \tfrac{1}{3}x^3$$

Hence $[T]_\beta^\alpha = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$.

Thus
$$[UT]_\beta = [U]_\alpha^\beta [T]_\beta^\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = [I]_\beta,$$ where $I: P_2(\mathbb{R}) \to P_2(\mathbb{R})$ is the identity transformation.

This confirms the fundamental theorem of calculus in a special case!

Definition The $n\times n$ identity matrix I_n is defined by $(I_n)_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i\neq j \end{cases}$.

Hence $I_1 = (1)$, $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, etc.

We summarize basic properties of matrix multiplication.

Theorem 2.12. Let $A \in M_{m\times n}(F)$, $B, C \in M_{n\times p}(F)$, and $D, E \in M_{q\times m}(F)$. Then

a) $A(B+C) = AB + AC$ and $(D+E)A = DA + EA$.

b) $a(AB) = (aA)B = A(aB)$ for any scalar $a \in F$.

c) $I_m A = A = A I_n$.

d) If $\dim(V) = n$ and $I: V \to V$ is the identity transformation, then $[I]_\beta = I_n$ for any ordered basis β for V.

Proof.

See textbook.

Compare to the basic properties of the composition of lin. transformations (Theorem 2.10).

Calculating value of a lin. transf. using its matrix representation

Theorem 2.14.

Let $T: V \to W$ be linear, V, W fin. dim. v.s.'s with ordered bases β and γ, respectively.
Then, for each $u \in V$ we have
$$[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta.$$

$m\times n$ matrix

$n\times 1$ matrix

vector in W

matrix multiplication

its coordinate vector, viewed as an $m\times 1$ matrix

Proof.

Suppose $\beta = \{v_1, \dots, v_n\}$, $\gamma = \{w_1, \dots, w_m\}$ – ordered bases for V and W, respectively.
Let $x \in V$, say $x = a_1 v_1 + \dots + a_n v_n$.
That is, $[x]_\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.

Let $B = [T]_\beta^\gamma$. Then
$$T(x) = a_1 T(v_1) + \dots + a_n T(v_n) = a_1\left(\sum_{i=1}^m B_{i1} w_i\right) + \dots + a_n\left(\sum_{i=1}^m B_{in} w_i\right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_j B_{ij}\right) w_i.$$
Hence
$$[T(x)]_\gamma = \begin{pmatrix} \sum_{j=1}^n a_j B_{1j} \\ \vdots \\ \sum_{j=1}^n a_j B_{mj} \end{pmatrix} = B\cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$ – as wanted.

Example. Let $T: P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be given by $T(f(x)) = f'(x)$.

Then
$$[T]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$ – calculated in a previous example
β, γ – standard ordered bases.

Let $p(x) \in P_3(\mathbb{R})$ be arbitrary, for example $p(x) = 2 - 4x + x^2 + 3x^3$.
Then $T(p(x)) = p'(x) = -4 + 2x + 9x^2$.
Hence:

$$[T(p(x))]_\gamma = [p'(x)]_\gamma = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

But also

$$[T]_\beta^\gamma \, [p(x)]_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -4 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}$$ — illustrating Theorem 2.14.

Associating a linear transformation to a matrix

Definition Let $A \in M_{m\times n}(F)$. We denote by L_A the mapping $L_A : F^n \to F^m$ defined by $L_A(x) = Ax$.

regarded as column vectors.

We call L_A a left-multiplication transformation.

Example.

Let $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \in M_{2\times 3}(\mathbb{R})$, hence $L_A : \mathbb{R}^3 \to \mathbb{R}^2$.

If $x = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}$ then $L_A(x) = Ax = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \end{pmatrix}$.

Theorem 2.15 (Properties of L_A)

Let $A \in M_{m\times n}(F)$. Then $L_A : F^n \to F^m$ is linear.

If $B \in M_{m\times n}(F)$ and β, γ are the standard ordered bases for F^n and F^m, resp., then:

a) $[L_A]_\beta^\gamma = A$.

b) $L_A = L_B \iff A = B$.

c) $L_{A+B} = L_A + L_B$, $L_{aA} = a \cdot L_A$ for all $a \in F$.

d) If $T : F^n \to F^m$ is lin., then there is a unique $C \in M_{m\times n}(F)$ s.t. $T = L_C$. In fact, $C = [T]_\beta^\gamma$.

e) If $E \in M_{n\times p}(F)$, then $L_{AE} = L_A L_E$.

f) If $m = n$, then $L_{I_n} = I_{F^n}$.

Proof. Linearity of L_A is clear by Theorem 2.12.

a) The j^{th} column of $[L_A]_\beta^\gamma$ is $L_A(e_j) = Ae_j$, which is also the j^{th} column of A.
So $[L_A]_\beta^\gamma = A$.

b) "$\Leftarrow$": clear
"$\Rightarrow$": If $L_A = L_B$, then by (a), $A = [L_A]_\beta^\gamma = [L_B]_\beta^\gamma = B$.

d) Let $T : F^n \to F^m$ be lin, let $C = [T]_\beta^\gamma$.
By Theorem 2.14,
$[T(x)]_\gamma = [T]_\beta^\gamma [x]_\beta$, or $T(x) = Cx = L_C(x)$ for all $x \in F^n$.
So $T = L_C$. The uniqueness of C follows from (b).

e) $(AE)e_j$ = the j^{th} column of AE $= A(Ee_j)$ — both equalities are easy to see by writing out the products.

Thus $L_{AE}(e_j) = (AE)e_j = A(Ee_j) = L_A(Ee_j) = L_A(L_E(e_j))$.

Hence $L_{AE} = L_A L_E$ (by the corollary to Theorem 2.6, if to linear transf.'s agree on a basis then they are equal).

(c), (f) — Exercise.

Theorem 2.16. (Matrix multiplication is associative)

Let $A \in M_{m\times n}(F)$, $B \in M_{n\times p}(F)$, $C \in M_{p\times r}(F)$. Then
$A(BC) = (AB)C$.

Proof.

We have (using Theorem 2.15 (e) and associativity of the composition of functions)
$L_{A(BC)} = L_A L_{BC} = L_A(L_B L_C) = (L_A L_B) L_C = L_{AB} L_C = L_{(AB)C}$.
By Theorem 2.15 (b), $A(BC) = (AB)C$.

Invertibility

Definition. Let V and W be v.s. and $T: V \to W$ linear.
A function $U: W \to V$ is an inverse of T if $TU = I_W$ and $UT = I_V$.
If T has an inverse, then T is invertible.
If T is invertible, then the inverse of T is unique and is denoted by T^{-1}.

Basic facts about invertible functions.
1) Let T and U be invertible. Then the following holds:
a) $(TU)^{-1} = U^{-1}T^{-1}$
b) $(T^{-1})^{-1} = T$; in particular, T^{-1} is invertible.
2) T is invertible $\iff$ T is a bijection.

Proof. 2) "$\Rightarrow$" For any $y \in W$, $TT^{-1}(y) = I_W(y) = y$. Hence $y = T(\underbrace{T^{-1}(y)}_{\in V})$, so T is surjective.
Assume $T(x_1) = T(x_2)$, then $T^{-1}(T(x_1)) = T^{-1}(T(x_2))$, hence $x_1 = x_2$ — so T is injective.

Theorem 2.17. Let V, W be v.s., let $T: V \to W$ be lin. and invertible.
Then $T^{-1}: W \to V$ is also linear.

Proof.
Let $y_1, y_2 \in W$ and $c \in F$. Since T is both surjective and injective, there exist unique vectors $x_1, x_2 \in V$ s.t. $T(x_1) = y_1$ and $T(x_2) = y_2$.
Thus $x_1 = T^{-1}(y_1)$ and $x_2 = T^{-1}(y_2)$. And so
$T^{-1}(cy_1 + y_2) = T^{-1}(cT(x_1) + T(x_2)) = T^{-1}(T(cx_1 + x_2)) = I_V(cx_1 + x_2) = cx_1 + x_2 = cT^{-1}(y_1) + T^{-1}(y_2)$.

Example. Let $T: P_1(\mathbb{R}) \to \mathbb{R}^2$ be the lin. transf. defined by $T(a+bx) = (a, a+b)$.
Then $T^{-1}: \mathbb{R}^2 \to P_1(\mathbb{R})$ is defined by $T^{-1}(c,d) = c + (d-c)x$ — also linear, as Theorem 2.17 predicts.

• Recall the analogy between linear transformations and matrices.

Definition. Let $A \in M_{n\times n}(F)$. Then A is invertible if there exists $B \in M_{n \times n}(F)$ s.t. $AB = BA = I$.

Note. If A is invertible, then the matrix B such that $AB = BA = I$ is unique, called the inverse of A and denoted A^{-1}.
(If C were another such matrix, then $C = CI = C(AB) = (CA)B = IB = B$).

Example. The inverse of $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. Indeed, $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

Lemma. Let $T: V \to W$ be lin. and invertible, and $\dim(V) < \infty$. Then $\dim(V) = \dim(W)$.

Proof. Let $\beta = \{x_1, \ldots, x_n\}$ be a basis for V.
By Theorem 2.2, $\mathrm{Span}(T(\beta)) = R(T) = W$.
Next, T is a bijection, so:
$\dim(N(T)) = 0$ (as $N(T) = \{0\}$ as T is injective).
$\dim(R(T)) = \dim(W)$ (as $R(T) = W$).
Hence, by the dimension theorem, $\dim(V) = \dim(N(T)) + \dim(R(T)) = \dim(W)$.

Theorem 2.18 Let V, W be fin. dim. v.s. with ordered bases β and γ, resp.
Let $T: V \to W$ be lin.
Then T is invertible $\iff$ $[T]_\beta^\gamma$ is invertible.
Furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Proof.

"$\Rightarrow$" Suppose T is invertible.
By the Lemma, $\dim(V) = \dim(W) = n$. So $[T]_\beta^\gamma \in M_{n\times n}(F)$.
By definition, $T^{-1}: W \to V$ satisfies $TT^{-1} = I_W$ and $T^{-1}T = I_V$. Thus
$$I_n = [I_V]_\beta = [T^{-1}T]_\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma.$$
Similarly,
$$[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = I_n.$$
So $[T]_\beta^\gamma$ is invertible and $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$.

"$\Leftarrow$" Suppose $A = [T]_\beta^\gamma$ is invertible. Then there exists $B \in M_{n\times n}(F)$ s.t. $AB = BA = I_n$.
By Theorem 2.6, there exists $U \in \mathcal{L}(W,V)$ s.t.
$$U(w_j) = \sum_{i=1}^n B_{ij} v_i \quad \text{for } j = 1, \dots, n,$$
where $\gamma = \{w_1, \dots, w_n\}$, $\beta = \{v_1, \dots, v_n\}$.
It follows that $[U]_\gamma^\beta = B$.
To show that $U = T^{-1}$, notice that
$$[UT]_\beta = [U]_\gamma^\beta [T]_\beta^\gamma = BA = I_n = [I_V]_\beta \quad \text{— by Theorem 2.11.}$$
So $UT = I_V$, and similarly, $TU = I_W$.

Example. Let β and γ be the standard ordered bases of $P_1(\mathbb{R})$ and $\mathbb{R}^2$, resp.
For T given by $T(a+bx) = (a, a+b)$ from the previous example, we have
$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad [T^{-1}]_\gamma^\beta = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$
We have already checked that each of these matrices is the inverse of the other.

Corollary. Let $A \in M_{n\times n}(F)$. Then A is invertible $\iff$ L_A is invertible. Moreover, $(L_A)^{-1} = L_{A^{-1}}$.

Isomorphisms.

Sometimes two vector spaces may consist of objects of very different nature, but behave identically from the algebraic point of view. We describe a precise way of "identifying" vector spaces with each other.

Definition. Let V, W be v.s. We say that V is isomorphic to W if there exists a lin. transf. $T: V \to W$ that is invertible.
Such a lin. transf. is called an isomorphism from V onto W.

Note. 1) V is isomorphic to V (using I_V).
2) V is isomorphic to $W \iff W$ is isomorphic to V.
3) If V is isomorphic to W and W is isomorphic to Z, then V is isomorphic to Z. Exersice.

Thus isomorphism is an equivalence relation on vector spaces.

Example. Let $T: F^2 \to P_1(F)$ be given by $T(a_1, a_2) = a_1 + a_2 x$.
Then T is an isomorphism, so F^2 is isomorphic to $P_1(F)$.

Theorem 2.19. Let V, W be fin. dim. v.s. over F.
Then V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

Proof. "$\Rightarrow$" Let $T: V \to W$ be an isomorphism from V to W.
By the lemma above, $\dim(V) = \dim(W)$.

"$\Leftarrow$" Suppose $\dim(V) = \dim(W)$, and let $\beta = \{v_1, \dots, v_n\}$, $\gamma = \{w_1, \dots, w_n\}$ be bases for V and W, resp.
By Theorem 2.6, there exists $T: V \to W$ s.t. T is lin. and $T(v_i) = w_i$ for $i = 1, \dots, n$.
By Theorem 2.2,
$R(T) = \mathrm{Span}(T(\beta)) = \mathrm{Span}(\gamma) = W$, so T is surjective.
By Theorem 2.5, T is also injective.
Hence T is an isomorphism.

Corollary. Let V be a v.s. over F.
Then V is isomorphic to F^n if and only if $\dim(V) = n$.

Up to this point, we have associated linear transformations with their matrix representations, and we have seen many analogies between the operations on $\mathcal{L}(V,W)$ and $M_{m\times n}(F)$.
Now we can show that these two spaces may be identified.

Theorem 2.20.
Let V, W be v.s. over F, $\dim(V) = n$, $\dim(W) = m$.
Let β, γ be ordered bases for V and W, respectively.
Then the function $\Phi: \mathcal{L}(V,W) \to M_{m\times n}(F)$ defined by
$$\Phi(T) = [T]_\beta^\gamma \quad \text{for all } T \in \mathcal{L}(V,W)$$
is an isomorphism.

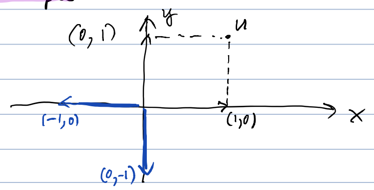
Proof.
By Theorem 2.8, Φ is linear. So remains to show Φ is a bijection.
That is, we need to show that for every $A \in M_{m\times n}(F)$, there is a unique lin. transf. $T: V \to W$ s.t. $\Phi(T) = A$.
Let $\beta = \{v_1, \dots, v_n\}$, $\gamma = \{w_1, \dots, w_m\}$, and let $A \in M_{m\times n}(F)$ be given.
By Theorem 2.6, there exists a unique lin. transf. $T: V \to W$ s.t.
$$T(v_j) = \sum_{i=1}^{m} A_{ij} w_i \quad \text{for } 1 \le j \le n.$$
But this means that $[T]_\beta^\gamma = A$, or $\Phi(T) = A$. Thus Φ is an isomorphism.

Corollary. If $\dim(V)=n$, $\dim(W)=m$, then $\dim(\mathcal{L}(V,W))=mn$.
(by the previous theorem, as $\dim(M_{m\times n}(F))=mn$).

Change of coordinate matrix

We have seen that once we fix an ordered basis β of a v.s. V to every vector $u\in V$ we can assign its coordinates $[u]_\beta$. And similarly, for $T: V\to V$, we assign its matrix rep. $[T]_\beta$.
However, these coordinates depend on β! And can be different for another choice of an ordered basis.

Example.



$V=\mathbb{R}^2$

$\beta=\{(1,0),(0,1)\}$ – ordered basis

$[u]_\beta=\begin{pmatrix}1\\1\end{pmatrix}$

$\gamma=\{(-1,0),(0,-1)\}$ – another ordered basis for V.

$[u]_\gamma=\begin{pmatrix}-1\\-1\end{pmatrix}$.

We would like a method to calculate $[u]_\gamma$ from $[u]_\beta$, for an arbitrary choice of β and γ.

Definition. Let β and β' be two ordered bases for a fin. dim. v.s. V.
We define the change of coordinate matrix (or "change of basis matrix") to be $Q=[I_V]_{\beta'}^{\beta}$.

Theorem 2.22.
1) Q is invertible. (and $Q^{-1}=[I_V]_{\beta}^{\beta'}$).
2) For any $v\in V$, $[v]_\beta=Q[v]_{\beta'}$.

Proof.
1) As I_V is invertible, Q is also invertible by Thm 2.18
2) For any $v\in V$,
$[v]_\beta=[I_V(v)]_\beta=[I_V]_{\beta'}^{\beta}[v]_{\beta'}=Q[v]_{\beta'}$, by Theorem 2.14.

So, multiplying by Q changes the β'-coordinates of a vector into its β-coordinates.
And multiplying by Q^{-1} changes β-coordinates into β'-coordinates.

Example.
In the example above, $[I_V]_\gamma^\beta=\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$.
And $[u]_\gamma=\begin{pmatrix}-1\\-1\end{pmatrix}$.
Hence $[u]_\beta=\begin{pmatrix}-1&0\\0&-1\end{pmatrix}\begin{pmatrix}-1\\-1\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}$.

Definition A lin. transf. $T: V\to V$ from a v.s. V to itself is called a linear operator on V.
• Now we determine how to calculate $[T]_\beta$ from $[T]_{\beta'}$, for β,β' two ordered bases for V.

Theorem 2.23. Let T be a lin. operator on a fin. dim. v.s. V.
Let β,β' be ordered bases for V.
Let $Q=[I_V]_{\beta'}^{\beta}$ be the change of coordinate matrix, changing β'-coord's into β-coords.
Then
$[T]_{\beta'}=Q^{-1}[T]_\beta Q$.

Proof.
Recall that $T=I_VT=TI_V$.
$Q[T]_{\beta'}=[I_V]_{\beta'}^{\beta}[T]_{\beta'}^{\beta'}=[I_VT]_{\beta'}^{\beta}=[TI_V]_{\beta'}^{\beta}=[T]_\beta^\beta[I_V]_{\beta'}^{\beta}=[T]_\beta Q$. (by Theorem 2.11).
Therefore
$[T]_{\beta'}=Q^{-1}[T]_\beta Q$.

Example.

Consider the lin. operator T on $\mathbb{R}^2$ defined by $T(x,y) = (x+y,\ x-y)$.

Let $\beta = \{(1,0),(0,1)\}$ and $\beta' = \{(-1,0),(0,-1)\}$ be ordered bases.

By the previous example:

$Q = [I_V]_{\beta'}^{\beta} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. Also $Q^{-1} = [I_V]_{\beta}^{\beta'} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.

Also $[T]_\beta = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Hence $[T]_{\beta'} = Q^{-1}[T]_\beta Q = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Determinants

Definition

Let $A \in M_{n\times n}(F)$.

1) For any $1 \le i,j \le n$ we define the cofactor matrix of the entry of A in row i and column j to be the matrix $\tilde{A}_{ij} \in M_{(n-1)\times(n-1)}(F)$ obtained from A by deleting row i and column j.

2) The determinant of A, denoted $\det(A)$, is a scalar in F defined recursively as follows:
   - if $n=1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$.
   - for $n \ge 2$, we define
   $$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}) \quad \text{for any } 1 \le i \le n$$
   (this formula gives the same value for any i! See Theorem 4.4).

3) Equivalently, we have
   $$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij}) \quad \text{for any } 1 \le j \le n.$$

Example. Let's consider the case $n=2$.

Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \in M_{2\times 2}(F)$ be given.

According to the definition, we can evaluate its determinant along any row i.

Let's take $i=1$.

Then the cofactor matrices are $\tilde{A}_{1,1} = (A_{22})$ and $\tilde{A}_{1,2} = (A_{21})$.

So $\det(\tilde{A}_{1,1}) = A_{22}$, $\det(\tilde{A}_{1,2}) = A_{21}$ and

$$\det(A) = \sum_{j=1}^{2} (-1)^{1+j} A_{1j} \cdot \det(\tilde{A}_{1,j}) = A_{11} \cdot A_{22} - A_{12} A_{21}. \quad \text{— the familiar formula.}$$

Example.

Let $A = \begin{pmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$.

Again, let's calculate $\det(A)$ using cofactors along the 1^{st} row. We obtain:

$$\det(A) = (-1)^{1+1} A_{11} \det(\tilde{A}_{1,1}) + (-1)^{1+2} A_{12} \cdot \det(\tilde{A}_{1,2}) + (-1)^{1+3} A_{13} \det(\tilde{A}_{1,3}) =$$

$$= (-1)^2 \cdot 1 \cdot \det\begin{pmatrix} -5 & 2 \\ 4 & -6 \end{pmatrix} + (-1)^3 \cdot 3 \cdot \det\begin{pmatrix} -3 & 2 \\ -4 & -6 \end{pmatrix} + (-1)^4 \cdot (-3) \cdot \det\begin{pmatrix} -3 & -5 \\ -4 & 4 \end{pmatrix} =$$

$$= 1 \cdot (-5 \cdot (-6) - 2 \cdot 4) - 3(-3 \cdot (-6) - 2 \cdot (-4)) - 3(-3 \cdot 4 - (-5) \cdot (-4)) =$$

$$= 1 \cdot 22 - 3 \cdot 26 - 3 \cdot (-32) = 40.$$

Properties of the determinant (See Sections 4.2-4.4 in the textbook for the proofs)

Let $A \in M_{n\times n}(F)$. If B is a matrix obtained from A by

1) switching two rows (or two columns), then
$\det(B) = -\det(A)$.
2) multiplying a row (or a column) of A by a scalar $c \in F$, then
$\det(B) = c\cdot\det(A)$.
3) adding a multiple of row i to row j (or a multiple of column i to column j), then
$\det(B) = \det(A)$.

These properties are helpful for computing determinants.

We also have the following properties:

4) If $B \in M_{n\times n}(F)$, then
$\det(AB) = \det(A)\cdot\det(B)$,
5) A is invertible if and only if $\det(A) \neq 0$. Furthermore,
$$\det(A^{-1}) = \frac{1}{\det(A)}.$$
6) If $I_n \in M_{n\times n}(F)$ is the identity matrix, then
$\det(I_n) = 1$.
7) $\det(A) = \det(A^t)$.

The operations on the rows of a matrix described in 1), 2) and 3) above are called elementary row operations.

Fact. Using these operations, we can transform any square matrix into an upper triangular matrix. That is, a matrix of the form

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ & A_{22} & \cdots & A_{2n} \\ 0 & & \ddots & \\ & & & A_{nn} \end{pmatrix}$$ – all entries below the diagonal are 0.

Fact. If $A \in M_{n\times n}(F)$ is upper triangular, then $\det(A) = A_{11}\cdot A_{22}\cdot\ \cdots\ \cdot A_{nn}$.

These two facts simplify calculating the determinants.

Example.

Let $B = \begin{pmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 4 & -4 & 4 \end{pmatrix}$.

Applying elementary rows operations, we have

$$B \xrightarrow{(1)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 4 & -4 & 4 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 0 & -10 & -6 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 24 \end{pmatrix}.$$

exchanging rows 1 and 2 adding $2\times$(row 1) to row 3 adding $10\times$(row 2) to row 3

As (3) doesn't change the determinant, we have

$\det\begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 4 & -4 & 4 \end{pmatrix} = -2\cdot 1\cdot(24) = -48$, and as (1) only changes the sign of det, we have $\det(B) = 48$.

Eigenvalues and eigenvectors.

Definition. 1) A lin. operator T on a fin. dim. v.s. V is called diagonalizable if there is an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix. That is,

$$[T]_\beta = \begin{pmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{nn} \end{pmatrix} \quad \text{for some } A_{11}, \dots, A_{nn} \in F.$$

2) A matrix $A \in M_{n\times n}(F)$ is diagonalizable if A is similar to a diagonal matrix.

Recall: two matrices $A, B \in M_{n\times n}(F)$ are similar if there is an invertible matrix $Q \in M_{n\times n}(F)$ such that $B = Q^{-1}AQ$.

Theorem. Let $T: V \to V$ be a lin. operator, $\dim(V) < \infty$ and β, γ ordered bases for V.
Then $\det([T]_\beta) = \det([T]_\gamma)$.

Proof.
There exists an invertible matrix Q s.t. $[T]_\gamma = Q^{-1}[T]_\beta Q$ (namely, the change of coordinates matrix $[I_V]_\gamma^\beta$ converting γ-coordinates to β-coordinates).
Then, using the basic properties of det, we have:

$$\det([T]_\gamma) = \det(Q^{-1}[T]_\beta Q) = \det(Q^{-1})\cdot\det([T]_\beta)\cdot\det(Q) = (\det Q)^{-1}\cdot\det Q\cdot\det([T]_\beta) = $$
$$= \det([T]_\beta).$$

Definition. For a lin. operator T, we define its determinant, $\det T$, as follows:
choose any ordered basis β for V and take $\det T = \det([T]_\beta)$.
(by the previous theorem, the choice of β doesn't matter).

Proposition
a) T is bijective $\Leftrightarrow$ $\det T \neq 0$.
b) T is bijective $\Rightarrow$ $\det(T^{-1}) = (\det T)^{-1}$.
c) If $U: V \to V$ is another lin. operator on V, then $\det(TU) = \det T \cdot \det U$.

Proof
Exercise, follows from the analogous properties of the matrix determinant.

Theorem. Let $T: V\to V$ be a lin. operator, $\dim(V) < \infty$, β an ordered basis for V. Then:
T is diagonalizable $\Leftrightarrow$ $[T]_\beta$ is a diagonalizable matrix.

Proof.
Let $\beta = \{v_1, \dots, v_n\}$.
$\Rightarrow$ Assume that T is diagonalizable. This means that there is an ordered basis γ for V such that $D = [T]_\gamma$ is a diagonal matrix. Let $[I_V]_\gamma^\beta$ be the change of coordinates matrix. Then $[T]_\gamma = Q^{-1}[T]_\beta Q$, so $[T]_\gamma$ and $[T]_\beta$ are similar, so $[T]_\beta$ is diagonalizable.
$\Leftarrow$ Exercise.

Corollary. $A \in M_{n\times n}(F)$ is diagonalizable $\Leftrightarrow$ L_A is diagonalizable.

Problem. When is A/T diagonalizable?

Theorem. T is diagonalizable $\Leftrightarrow$ there is an ordered basis $\beta = \{v_1, \dots, v_n\}$ for V and scalars $\lambda_1, \dots, \lambda_n \in F$ such that
$T(v_j) = \lambda_j v_j$ for $1 \le j \le n$.

Proof.
If $D = [T]_\beta$ is a diagonal matrix, then for each vector $v_j \in \beta$ we have
$T(v_j) = \sum_{i=1}^n D_{ij} v_i = D_{jj} v_j = \lambda_j v_j$, where $\lambda_j = D_{jj}$.
Conversely, if β is an ord. basis for V s.t. $T(v_j) = \lambda_j v_j$, then

$$[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

This argument motivates the following definition.

Definition.
1) A non-zero vector $v \in V$ is an **eigenvector** of T if $T(v) = \lambda v$ for some $\lambda \in F$. We call λ the **eigenvalue** of T corresponding to the eigenvector v.
2) Let $A \in M_{n\times n}(F)$. A non-zero $v \in F^n$ is an **eigenvector** of A if $Av = \lambda v$ for some $\lambda \in F$.
   And λ is the **eigenvalue** of A corresponding to the eigenvector v.
3) The elements in a basis β as in the last theorem are eigenvectors, and the λ_i's are the respective eigenvalues.

Theorem 5.2. A scalar $\lambda \in F$ is an eigenvalue of $T \iff \det(T - \lambda I_V) = 0$

Proof. We have
$\lambda \in F$ is an eigenvalue of $T \iff T(v) = \lambda v$ for some $v \neq 0$ in $V \iff \overbrace{(T - \lambda I_V)}^{\text{lin. operator on } V}(v) = 0$ for some $v \neq 0$ in $V \iff N(T - \lambda I_V) \neq \{0\} \overset{\text{Thm 2.4, 2.5}}{\iff} T - \lambda I_V$ is not bijective $\underset{\text{properties of det.}}{\iff} \det(T - \lambda I_V) = 0$.

Corollary. Let $A \in M_{n\times n}(F)$. Then $\lambda \in F$ is an eigenvalue of $A \iff \det(A - \lambda I_n) = 0$.

Example. Let $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$.
Then $\det(A - \lambda I_2) = \det\begin{pmatrix} 1-\lambda & 1 \\ 4 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 4 = (\lambda - 3)(\lambda + 1)$.
Hence by the corollary, the eigenvalues of A are the solutions to $(\lambda-3)(\lambda+1) = 0$ — which are $3, -1$.

Definition 1) The polynomial $f(t) = \det(A - tI_n)$ in the variable t is called the **characteristic polynomial** of A.
2) Given a lin. operator $T: V \to V$, $\dim(V) < \infty$, and β an ordered basis for V, we define the **characteristic polynomial** of T to be the char. polynomial of $A = [T]_\beta$:
$$f(t) = \det(A - tI).$$

Note. Similar matrices have the same char. polynomial, so is well defined.

Properties of char. polynomial.
Let $A \in M_{n\times n}(F)$ be given, and let $f(t)$ be its char. polynomial.
1) $f(t)$ is a polynomial of **degree n** with leading coefficient $(-1)^n$:
   $f(t) = (-1)^n t^n + c_{n-1} t^{n-1} + \dots + c_0$ for some $c_0, \dots, c_{n-1} \in F$.
2) A scalar $\lambda \in F$ is an eigenvalue of $A \iff f(\lambda) = 0$.
3) A has at most n distinct eigenvalues (as $f(t)$ has at most n roots).
4) If $\lambda \in F$ is an eigenvalue of A, then a vector $x \in F^n$ is an eigenvector of A corresp. to $\lambda \iff x \neq 0$ and $x \in N(L_A - \lambda I_{F^n})$.

Example. Let's consider $A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$ again, and let's find it
1) The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$.
2) Let $B_1 = A - \lambda_1 I_2 = \begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}$.
   Then $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ is an eigenvector of A corresponding to λ_1 — by (4) above.
   $\iff x \neq 0$ and $x \in N(L_{B_1}) \iff x \neq 0$ and $\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. $\iff$

$\iff \begin{pmatrix} -2x_1+x_2 \\ 4x_1-2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

The set of all solutions to this system of lin. equations is
$\left\{ t\begin{pmatrix} 1 \\ 2 \end{pmatrix} : t\in\mathbb{R} \right\}$.

Hence $x\in\mathbb{R}^2$ is an eigenvector corresp. to $\lambda_1=3 \iff x = t\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ for some $t\neq 0$.

3) Let $B_2 = A-\lambda_1 I_2 = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$. Hence:
$x\in\mathbb{R}^2$ is an e.vec. of A corresp. to $\lambda_2 \iff x\neq 0$ and $x\in N(L_{B_2}) \iff B_2\cdot x = 0 \iff \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
$\iff \begin{cases} 2x_1+x_2=0 \\ 4x_1+2x_2=0 \end{cases}$.

Hence $N(L_{B_2}) = \left\{ t\begin{pmatrix} 1 \\ -2 \end{pmatrix} : t\in\mathbb{R} \right\}$. Thus x is an e.vec. corresp. to $\lambda_2=-1 \iff x = t\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ for some $t\neq 0$.

Notice that $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^2$ consisting of e.vectors of A. Thus L_A, and hence A, is diagonalizable.

Determining eigenvectors and eigenvalues of a lin. operator

Let V be a v.s., $\dim(V)=n$. Let β be an ordered basis for V.
Let $T\in\mathcal{L}(V)$ be a lin. operator on V.
Summarizing the results of the previous section, we describe how to determine the e.val's and the e.vec's of T.

1) Determine the matrix representation $[T]_\beta$ of T.

2) Determine the e.val's of T.
$\lambda\in F$ is an e.val of $T \iff \lambda$ is a root of the char. polynomial of T.
That is, we need to find the solutions $x\in F$ of $\det([T]_\beta - xI_n)=0$.
There are at most n distinct solutions $\lambda_1,\dots,\lambda_n$.

3) Now for each e.val. λ of T, we can determine the corresponding e.vec's. We have:
$T(v)=\lambda v \iff (T-\lambda I_V)(v)=0 \iff [T-\lambda I_V]_\beta [v]_\beta = 0$.
Therefore, eigenvectors corresponding to λ are the solutions of this system of linear equations. (more precisely, solving this system we find the β-coordinates $[v]_\beta$, which then determines v).

Theorem 5.5

Let $T\in\mathcal{L}(V)$ be a lin. operator on V, and let $\lambda_1,\dots,\lambda_k$ be distinct eigenvalues of T. If $v_1,\dots,v_k$ are e.vect's of T s.t. v_i corresponds to λ_i, then the set
$\{v_1,\dots,v_k\}$
is lin. indep.

Proof.
By induction on k.
$k=1$. As $v_1\neq 0$, $\{v_1\}$ is lin. indep.
Induction step, $k-1 \Rightarrow k$.
Suppose we know the theorem for $k-1$ distinct eigenvalues, and lets prove it for k.
Suppose $a_1,\dots,a_k\in F$ are such that
$a_1v_1+\dots+a_kv_k=0$.
Applying the linear transf. $T-\lambda_k I_V$ to both sides and using linearity, we get:
$(T-\lambda_k I_V) = T(0) - \lambda_k I_V(0) = 0-0=0$.
$(T-\lambda_k I_V)(a_1v_1+\dots+a_kv_k) = (a_1T(v_1)+\dots+a_kT(v_k)) - \lambda_k(a_1v_1+\dots+a_kv_k) \overset{\text{as } T(v_i)=\lambda_i v_i}{=} a_1\lambda_1v_1+\dots+a_k\lambda_kv_k - \lambda_k a_1v_1+\dots+\lambda_k a_k v_k = a_1(\lambda_1-\lambda_k)v_1+\dots+a_{k-1}(\lambda_{k-1}-\lambda_k)v_{k-1} = 0$.
By Induction Hypothesis, $\{v_1,\dots,v_{k-1}\}$ are lin. indep., so
$a_1(\lambda_1-\lambda_k)=\dots=a_{k-1}(\lambda_{k-1}-\lambda_k)=0$.

Since $\lambda_1, \dots, \lambda_k$ are distinct by assumption, $\lambda_i - \lambda_k \neq 0$ for $i = 1, \dots, k-1$.
Thus $a_1 = \dots = a_{k-1} = 0$.
Hence $a_k v_k = 0$. But as $v_k \neq 0$ (as an eigenvector), we get $a_k = 0$.

Corollary Let $T \in \mathcal{L}(V)$ and $\dim(V) = n$.
If T has n distinct e.val's, then T is diagonalizable.
Proof.
Let $\lambda_1, \dots, \lambda_n$ be n distinct e.val's of T. For each i, let v_i be an eigenvector corresp. to λ_i. By the theorem, $\{v_1, \dots, v_n\}$ is lin. indep. Since $\dim(V) = n$, this set is a basis for V.
Thus V has a basis consisting of eigenvectors for T, so T is diagonalizable.

Ex The converse of Thm 5.5 is false.
For example, the identity operator I_V has only one eigenvalue, namely $\lambda = 1$. However it is diagonalizable!

Def A polynomial $f(t) \in P(F)$ splits over F if there are scalars $c, a_1, \dots, a_n \in F$ (not necessarily distinct) such that
$f(t) = c(t-a_1)(t-a_2) \cdot \dots \cdot (t-a_n)$.

Ex 1) $t^2 - 1 \in P_2(\mathbb{R})$ splits over $\mathbb{R}$, namely $t^2 - 1 = (t-1)(t+1)$.
2) $t^2 + 1 \in P_2(\mathbb{R})$ doesn't split over $\mathbb{R}$.
However, viewed as a polynomial in $P_2(\mathbb{C})$, it splits over $\mathbb{C}$: $t^2 + 1 = (t+i)(t-i)$.

Thm 5.6 The char. polynomial of any diagonalizable lin. operator splits.
Proof.
Let $n = \dim(V)$, $T \in \mathcal{L}(V)$ be diagonalizable, then there is an ordered basis β for V s.t. $[T]_\beta = D$, where D is of the form

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Let $f(t)$ be the char. polynomial of T. Then

$$f(t) = \det(D - tI) = \begin{pmatrix} \lambda_1 - t & & 0 \\ & \lambda_2 - t & \\ 0 & & \ddots \quad \lambda_n - t \end{pmatrix} = (\lambda_1 - t) \cdot \dots \cdot (\lambda_n - t) = (-1)^n (t - \lambda_1) \cdot \dots \cdot (t - \lambda_n).$$

Def Let λ be an e.val. of a lin. operator or matrix with char. polynomial $f(t)$. The (algebraic) multiplicity of λ is the largest positive integer k for which $(t-\lambda)^k$ is a factor of $f(t)$. (That is, $f(t)$ can be written as $f(t) = (t-\lambda)^k g(t)$ for some polynomial $g(t)$).

Ex Let $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{pmatrix}$, then $f(t) = -(t-3)^2(t-4)$. Hence $\lambda = 3$ is an e.val. of A with mult. 2, and $\lambda = 4$ is an e.val. of A with mult. 1.

Def Let $T \in \mathcal{L}(V)$, λ an eigenvalue of T. We define E_λ, the eigenspace of T corresp. to λ, as

$$E_\lambda = \{x \in V : T(x) = \lambda x\} = N(T - \lambda I_V).$$ (and similary for a matrix).

Note that this is a subspace of V, consisting of 0 and the e.vects of T corresp to λ.

Thm 5.7 Let $T \in \mathcal{L}(V)$, $\dim(V) < \infty$, λ an e.val. of T with multiplicity m.
Then $1 \le \dim(E_\lambda) \le m$.

Proof.
Choose an ordered basis $\{v_1, \dots, v_p\}$ for E_λ.
By the replacement thm, can extend it to an ordered basis $\beta = \{v_1, \dots, v_p, v_{p+1}, \dots, v_n\}$ for V.
Let $A = [T]_\beta$.
Notice that v_i, $i = 1, \dots, p$, is an eigenvector of T corresp to λ. Hence

$$A = \left(\begin{array}{c|c} \lambda I_p & B \\ \hline O & C \end{array}\right). \quad \text{Then}$$

$$f(t) = \det(A - tI_n) = \det \left(\begin{array}{c|c} (\lambda - t) I_p & B \\ \hline O & C - tI_{n-p} \end{array}\right) \overset{\text{exercise!}}{=} \det((\lambda - t) I_p) \det(C - tI_{n-p}) =$$
$$= (\lambda - t)^p g(t),$$

where $g(t)$ is a polynomial.
Thus $(\lambda - t)^p$ is a factor of $f(t)$, hence the mult. of λ is at least p. But $\dim(E_\lambda) = p$, so $\dim(E_\lambda) \le m$.

Lemma Let $T \in \mathcal{L}(V)$, $\lambda_1, \dots, \lambda_k$ distinct e.vals of T.
Let $v_i \in E_{\lambda_i}$ for each $i = 1, \dots, k$.
If $v_1 + \dots + v_k = 0$ <u>then</u> $v_i = 0$ for all i.

Pf Suppose otherwise, say we have $v_i \ne 0$ for $1 \le i \le m$, and $v_i = 0$ for $i > m$, for some $1 \le m \le k$.
Then for each $i \le m$, v_i is an e.vect of T corresp. to λ_i. (as $v_i \in E_{\lambda_i} \setminus \{0\}$),
and $v_1 + \dots + v_m = 0$.
But this contradicts Thm 5.5 as $v_1, \dots, v_m$ must be lin. indep. Therefor $v_i = 0$ for all $i = 1, \dots, k$.

Thm 5.8 Let $T \in \mathcal{L}(V)$, let $\lambda_1, \dots, \lambda_k$ be distinct e.vals of T.
For each $i = 1, \dots, k$, let S_i be a finite lin. indep. subset of E_{λ_i}.
Then $S = S_1 \cup \dots \cup S_k$ is also a lin. indep. subset of V.

Proof.
Suppose that $S_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,n_i}\}$ for each $i = 1, \dots, k$.
Then $S = \{v_{i,j} : 1 \le j \le n_i,\ 1 \le i \le k\}$.
Let $\{a_{i,j}\}$ be any scalars in F s.t.

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} v_{ij} = 0.$$

For each i, let $w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}$. Then: $w_i \in \overset{\text{subspace!}}{E_{\lambda_i}}$, and $w_1 + \dots + w_k = 0$.
By the lemma, $w_i = 0$ for all $i = 1, \dots, k$.
But as each S_i is indep., it follow that $a_{ij} = 0$ for all j.
Hence S is lin. indep.

Thm 5.9 Let $T \in \mathcal{L}(V)$, $\dim(V) < \infty$, and assume that the char. poly. of T splits.
Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T. Then
a) T is diagonalizable $\iff$ the multiplicity of λ_i is equal to $\dim(E_{\lambda_i})$ for all i.
b) If T is diagonalizable and β_i is an ordered basis for E_{λ_i} for each i, then
$\beta = \beta_1 \cup \dots \cup \beta_k$ is an ordered basis for V consisting of e.vects of T.

Proof.
For each i, let m_i denote the multiplicity of λ_i, $d_i = \dim(E_{\lambda_i})$, and $n = \dim(V)$.
$\Rightarrow$ Suppose that T is diagonalizable.
Let β be a basis for V consisting of e.vects of T.
For each i, let $\beta_i = \beta \cap E_{\lambda_i}$.
Let $n_i = |\beta_i|$.
Then:

- $n_i \le d_i$ for each i (because β_i is a lin. indep. subset of the subspace E_{λ_i} and $\dim(E_{\lambda_i}) = d_i$).
- $d_i \le m_i$ (by Thm 5.7).
- $\sum_{i=1}^{k} n_i = n$ (because β contains n vectors).
- $\sum_{i=1}^{k} m_i = n$ (because the degree of the char. poly. of T is equal to the sum of the mult. of the eigenvalues, on the one hand, and is equal to $\dim(V) = n$ on the other hand).

Thus:
$$n = \sum_{i=1}^{k} n_i \le \sum_{i=1}^{k} d_i \le \sum_{i=1}^{k} m_i = n.$$
It follows that
$$\sum_{i=1}^{k} (m_i - d_i) = 0.$$
Since $(m_i - d_i) \ge 0$ for all i, we conclude that $m_i = d_i$ for all i.

$\Leftarrow$ Conversely, suppose that $m_i = d_i$ for all i.
For each i, let β_i be an ordered basis for E_{λ_i}, and let $\beta = \beta_1 \cup \dots \cup \beta_k$.
By Thm 5.8, β is lin. indep.
Furthermore, since $d_i = m_i$ for all i by assumption, β contains
$$\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} m_i = n$$
vectors.

Therefore β is an ordered basis for V consisting of e.vects of V. Hence T is diagz.

This theorem concludes our study of the diagonalization problem. Let's summarize.

Test for diagonalization
Let T be a lin. operator on an n-dim. v.s. V.
Then T is diagonalizable if and only if both of the following conditions hold.
1) The char. polynomial of T splits.
2) For each e.val λ of T, the multiplicity of λ equals $\dim E_\lambda = \dim N(T - \lambda I_V) =$
$= n - \operatorname{rank}(T - \lambda I_V)$.

The same conditions can be used to test if a square matrix is diagz, because A is diagz $\iff$ the operator L_A is diagz.

Example
Let $A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$, and we test its diagonalizability.
The char. poly. $f(t) = \det(A - tI_3) = \det \begin{pmatrix} 3-t & 1 & 0 \\ 0 & 3-t & 0 \\ 0 & 0 & 4-t \end{pmatrix} = (4-t)(3-t)^2 = -(t-4)(t-3)^2$.
This shows that $f(t)$ splits, so condition (1) for diagz. holds.

E. vals:
$\lambda_1 = 4$ — mult. 1
$\lambda_2 = 3$ — mult. 2
Condition (2) is automatically satisfied for λ_1 (as by Thm 5.7, $1 \le \dim(E_{\lambda_1}) \le \text{mult}\,\lambda_1 = 1$)
So only need to check (2) for λ_2.
The matrix
$A - \lambda_2 I_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ has rank 2 (the rank of $L_{A-\lambda_2 I_3}$, equivalently the max. number of lin. indep. columns).
$\dim E_{\lambda_2} = 3 - \text{rank}(A - \lambda_2 I_3) = 3 - 2 = 1 \ne 2$, the mult. of λ_2.
Hence A is not diagz.

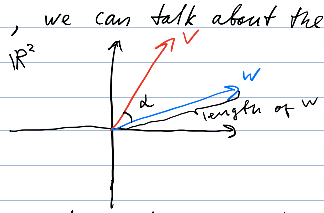
Example.
Let $A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$.
$f(t) = \det(A - tI_2) = (t-1)(t-2)$.
Hence $\lambda_1 = 1$, $\lambda_2 = 2$ are the e. vals, both of mult. 1. Thus both conditions (1), (2) are satisfied and T is diagz.
$E_{\lambda_1} = N(L_A - 1 \cdot I_2) = \langle \begin{pmatrix} -2 \\ 1 \end{pmatrix} \rangle$.
$E_{\lambda_2} = N(L_A - 2 \cdot I_2) = \langle \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rangle$.

Hence $\beta_1 = \{\begin{pmatrix} -2 \\ 1 \end{pmatrix}\}$ is a basis for E_{λ_1}, and $\beta_2 = \{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\}$ is a basis for E_{λ_2}.
By the theorem $\beta = \beta_1 \cup \beta_2 = \{\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\}$ is a basis for $V = \mathbb{R}^2$ consisting of e. vects. So $[L_A]_\beta$ is a diag. matrix.
Let $Q = \begin{pmatrix} -2 & -1 \\ 1 & 1 \end{pmatrix}$.
Then $Q^{-1} A Q = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ — diagonal.

Fact. Let $A \in M_{n \times n}(F)$, let $\gamma = \{v_1, \dots, v_n\}$ be an ord. basis for F^n. Then
$[L_A]_\gamma = Q^{-1} A Q$, where
$$Q = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix}.$$

Inner products and norms

- In $V = \mathbb{R}^2$, we can talk about the length of a vector, the angle between two vectors, two vectors being orthogonal, etc.



- In a general v.s. V, these notions are not defined. For example, if $V = P_2(\mathbb{R})$, what is the length of a polynomial $3x^2 + 2x - 1$?
- In order to study these notions in general, we introduce an "upgraded" version of vector spaces.

Def
Let V be a v.s. over F (for $F = \mathbb{R}$ or $F = \mathbb{C}$).
An inner product on V is a function that assigns, to every ordered pair of vectors x and y in V, a scalar in F, denoted by $\langle x, y \rangle$, such that for all $x, y, z \in V$ and $c \in F$ the following holds:
a) $\langle x+z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$
b) $\langle cx, y \rangle = c \langle x, y \rangle$
} – linearity in the first variable
c) $\overline{\langle x, y \rangle} = \langle y, x \rangle$ (where "$\overline{}$" denotes complex conjugation). – conjugate symmetry
d) $\langle x, x \rangle > 0$ for $x \ne 0$ – positivity

Remark 1) If $F=\mathbb{R}$, then (c) reduces to $\langle x,y\rangle = \langle y,x\rangle$.
2) It follows from the definition that if $a_1, \dots, a_n \in F$ and $y, v_1, \dots, v_n \in V$, then

$$\left\langle \sum_{i=1}^{n} a_i v_i \, , \, y \right\rangle = \sum_{i=1}^{n} a_i \langle v_i, y\rangle.$$

Example. We define the standard inner product on F^n.
For $x=(a_1,\dots,a_n)$, $y=(b_1,\dots,b_n)$ in F^n, define

$$\langle x,y\rangle = \sum_{i=1}^{n} a_i \overline{b_i}.$$

We can verify that $\langle \cdot,\cdot\rangle$ satisfies the conditions (a) through (d).
For example, if $z=(c_1,\dots,c_n)$, we have for (a)

$$\langle x+z, y\rangle = \sum_{i=1}^{n} (a_i+c_i)\overline{b_i} = \sum_{i=1}^{n} a_i\overline{b_i} + \sum_{i=1}^{n} c_i\overline{b_i} = \langle x,y\rangle + \langle z,y\rangle.$$

For example, for $x=(1+i,4)$ and $y=(2-3i, 4+5i)$ in $\mathbb{C}^2$, $\langle x,y\rangle = (1+i)(2+3i)+4\cdot(4-5i) = 15-15i$.

When $F=\mathbb{R}$ the conjugations are not needed, and $\langle x,y\rangle$ gives the dot product from 33A.

Example If $\langle x,y\rangle$ is any inner product on a v.s. V and $r>0$, we may define another inner product by the rule $\langle x,y\rangle' = r\langle x,y\rangle$. (If $r\le 0$, then (d) would not hold.)

Example
Let $V=C(\mathbb{R})$, the v.s. of real-valued continuous functions on $\mathbb{R}$.
For $f,g\in V$, define

$$\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt.$$

(a) and (b) hold by the basic properties of integration, for example for (a) we have
$\langle f_1+f_2, g\rangle = \int_0^1 (f_1(t)+f_2(t))\,g(t)\,dt = \int_0^1 f_1(t)g(t)\,dt + \int_0^1 f_2(t)g(t)\,dt = \langle f_1,g\rangle + \langle f_2,g\rangle$.
(c) is clear, and (d) is easy to verify as $\int_0^1 (f(t))^2\,dt > 0$ for any continuous $f\neq 0$.
Thus, $\langle\cdot,\cdot\rangle$ is an inner product on $C(\mathbb{R})$.
Note that similarly, $\langle f,g\rangle' = \int_{-1}^{1} f(t)g(t)\,dt$ gives another inner product on $C(\mathbb{R})$.

Example
Let $A\in M_{m\times n}(F)$. We define the conjugate transpose of A as the $n\times m$ matrix A^* s.t. $(A^*)_{ij} = \overline{A_{ji}}$.
When $F=\mathbb{R}$, then A^* is simply A^t.
For example, if $A=\begin{pmatrix} i & 1+2i \\ 2 & 3+4i\end{pmatrix}$, then $A^* = \begin{pmatrix} -i & 2 \\ 1-2i & 3-4i\end{pmatrix}$.

Consider now $V=M_{n\times n}(F)$, and define $\langle A,B\rangle = tr(B^*A)$ for $A,B\in V$.
This defines an inner product on V, called the Frobenius inner product.
(see page 331, Example 5 for a proof that (a)–(d) hold.)

Def A v.s. V over F endowed with a specific inner product is called an inner product space.
If $F=\mathbb{C}$, V is called a complex inner product space.
If $F=\mathbb{R}$, V is called a real inner product space.

Remark. 1) If a v.s. V has an inner product $\langle x,y\rangle$ and W is a subspace of V, then W is also an inner product space when the same function $\langle x,y\rangle$ is restricted to the vectors $x,y\in W$.
As $P_n(\mathbb{R})$ is a subspace of $C(\mathbb{R})$, it follows that $P_n(\mathbb{R})$ can be equipped with (many different inner products.

Thm 6.1 (basic properties of inner products).

Let V be an inner product space. Then for any $x, y, z \in V$ and $c \in F$ we have

a) $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$

b) $\langle x, cy \rangle = \bar{c} \langle x, y \rangle$

c) $\langle x, 0 \rangle = \langle 0, x \rangle = 0$

d) $\langle x, x \rangle = 0 \iff x = 0$

e) If $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in V$, then $y = z$.

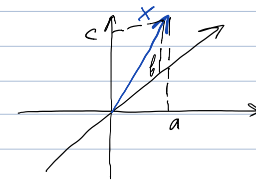
Proof.

(a) $\langle x, y+z \rangle = \overline{\langle y+z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} = \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle$.

(b) – (e). Exercize.

Example

Consider $\mathbb{R}^3$ with the standar inner product.
Then for $x = (a, b, c) \in \mathbb{R}^3$, the length of x is given by
$\sqrt{a^2+b^2+c^2} = \sqrt{\langle x, x\rangle}$.



By imitating what happens in $\mathbb{R}^3$, we can define length in an arbitrary inner product space.

Def Let V be an inner product space.
For any $x \in V$, we define the norm or length of x by $\|x\| = \sqrt{\langle x, x\rangle}$.

Example

Let $V = F^n$. If $x = (a_1, \dots, a_n)$, then
$\|x\| = \|(a_1, \dots, a_n)\| = \left(\sum_{i=1}^{n} |a_i|^2\right)^{\frac{1}{2}}$
is the Euclidean definition of length.

Many properties of the Euclidean length in $\mathbb{R}^3$ hold in general.

Thm 6.2

Let V be an inner product space over F. Then for all $x, y \in V$ and $c \in F$ we have:
a) $\|cx\| = |c| \cdot \|x\|$.
b) $\|x\| = 0 \iff x = 0$. (and $\|x\| \geq 0$ for any x).
c) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. (Cauchy – Schwarz Inequality)
d) $\|x+y\| \leq \|x\| + \|y\|$ (Triangle Inequality)

Proof.

c) If $y = 0$, then $\langle x, y\rangle = 0$ and $\|y\| = 0$, so the result holds.
Assume now $y \neq 0$.
For any $c \in F$ we have

$0 \leq \|x - cy\|^2 = \langle x - cy, x - cy\rangle = \langle x, x - cy\rangle - c\langle y, x - cy\rangle = \langle x, x\rangle - \bar{c}\langle x, y\rangle - c\langle y, x\rangle + c\bar{c}\langle y, y\rangle$.

In particular, if we take $c = \dfrac{\langle x, y\rangle}{\underbrace{\langle y, y\rangle}_{\neq 0 \text{ as } y \neq 0}}$, we have

We are using that $a \cdot \bar{a} = |a|^2$ and $a + \bar{a} = 2\,\mathrm{Re}(a)$ for any complex number a, and that $\langle x, y\rangle = \overline{\langle y, x\rangle}$

$$0 \leq \langle x, x\rangle - \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2},$$ and (c) follows.

d) We have

the real part of the complex number $\langle x, y\rangle$

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle = \|x\|^2 + 2\,\overbrace{\mathrm{Re}\,\langle x, y\rangle} + \|y\|^2 \leq$$
$$\leq \|x\|^2 + 2\,|\langle x, y\rangle| + \|y\|^2 \underset{\text{by (c)}}{\leq} \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Orthogonality

- As you may recall from earlier courses, for $\mathbb{R}^2$ and $\mathbb{R}^3$ there is another formula expressing the dot product of two vectors x and y:
  $\langle x, y\rangle = \|x\| \cdot \|y\| \cdot \cos\theta$
  where θ $(0 \leq \theta \leq \pi)$ is the angle between x and y.
  Notice also that non-zero vectors are perpendicular if and only if $\cos\theta = 0$, that is if and only if $\langle x, y\rangle = 0$.
- We generalize this to define perpendicularity in arbitrary inner product spaces.

Def. Let V be an inner product space.
1) Vectors x and y in V are orthogonal (perpendicular) if $\langle x,y\rangle = 0$.
2) A subset S of V is orthogonal if any two distinct vectors in S are orthogonal.
3) A vector x in V is a unit vector if $\|x\| = 1$.
4) A subset S of V is orthonormal if S is orthogonal and consists entirely of unit vectors.

Remark (Normalization)
1) $S = \{v_1, v_2, \ldots\}$ is orthonormal $\iff \begin{cases} \langle v_i, v_j\rangle = 0 & \text{for all } i \neq j \\ \langle v_i, v_i\rangle = 1 & \text{for all } i. \end{cases}$

2) If $S = \{v_1, v_2, \ldots\}$ is orthogonal, and $a_i \in F$ are any non-zero scalars, then the set $\{a_1v_1, a_2v_2, \ldots\}$ is also orthogonal (as $0 = \langle a_iv_i, a_jv_j\rangle = a_i\bar{a}_j\langle v_i, v_j\rangle \iff \langle v_i, v_j\rangle = 0$).

3) If x is any non-zero vector in V, then $y := \left(\frac{1}{\|x\|}\right)x$ is a unit vector. We say that y is obtained from x by normalizing.

4) In view of (2) and (3), we can obtain an orthonormal set from an orthogonal set by normalizing every vector in it.

Ex. In $\mathbb{R}^3$, $\{(1,1,0), (1,-1,1), (-1,1,2)\}$ is an orthogonal set of non-zero vectors, but it is not orthonormal.
Normalizing each of the vectors, we obtain an orthonormal set
$\left\{\frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{3}}(1,-1,1), \frac{1}{\sqrt{6}}(-1,1,2)\right\}$.

Orthonormal bases and Gram–Schmidt orthogonalization

Def Let V be an inner product space.
A subset S of V is an orthonormal basis for V if S is an ordered basis for V and S is orthonormal.

- Just as bases are the building blocks of vector spaces, orthonormal bases are the building blocks of inner product spaces.

Ex The standard ordered basis for F^n is an orthonormal basis for the inner product space F^n (with the standard inner product).

Ex The set $\left\{\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right), \left(\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}\right)\right\}$ is an orthonormal basis for $\mathbb{R}^2$.

Importance of orthonormal sets and bases is illustrated by the following theorem and its corollaries

Thm 6.3 Let V be an inner prod. space and $S = \{v_1, \ldots, v_k\}$ an orthogonal subset of V consisting of <u>non-zero</u> vectors.
If $y \in \mathrm{Span}(S)$, then $$y = \sum_{i=1}^{k} \frac{\langle y, v_i\rangle}{\|v_i\|^2} v_i.$$

Proof. Write $y = \sum_{i=1}^{k} a_iv_i$, where $a_1, \ldots, a_k \in F$. Then, for $1 \leq j \leq k$, we have
$$\langle y, v_j\rangle = \left\langle \sum_{i=1}^{k} a_iv_i, v_j\right\rangle = \sum_{i=1}^{k} a_i \underbrace{\langle v_i, v_j\rangle}_{=0 \text{ for all } i\neq j \text{ by orthogonality}} = a_j\langle v_j, v_j\rangle = a_j\|v_j\|^2.$$

So $a_j = \dfrac{\langle y, v_j\rangle}{\|v_j\|^2}$.

Corollary 1 If, in addition to the hypotheses of Thm. 6.3, S is orthonormal and $y \in \mathrm{Span}(S)$, then
$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i .$$

Corollary 2. Let V be an inner product space, and let S be an orthogonal subset of V consisting of non-zero vectors. Then S is lin. indep.

Proof.
Suppose that $v_1, \dots, v_k \in S$ and $\sum_{i=1}^{k} a_i v_i = 0$. As in the proof of Thm 6.3 with $y=0$, we have $a_j = \dfrac{\langle 0, v_j \rangle}{\|v_j\|^2} = 0$ for all j. So S is lin. indep.

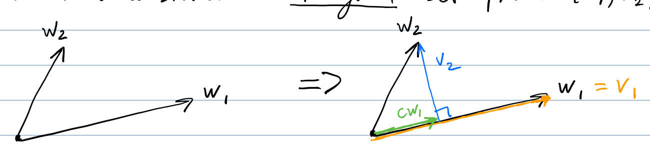
Ex By Corollary 2, the orthonormal set $\beta = \{ \underbrace{\tfrac{1}{\sqrt{2}}(1,1,0)}_{v_1}, \underbrace{\tfrac{1}{\sqrt{3}}(1,-1,1)}_{v_2}, \underbrace{\tfrac{1}{\sqrt{6}}(-1,1,2)}_{v_3} \}$ from a prev. example is an orthonormal basis for $\mathbb{R}^3$.
Let $x = (2,1,3)$. Using Corollary 1, it is easy to calculate the coordinates of x relatively to β:

$a_1 = \langle x, v_1 \rangle = 2 \cdot \frac{1}{\sqrt{2}} + 1 \cdot \frac{1}{\sqrt{2}} + 3 \cdot 0 = \frac{3}{\sqrt{2}}$, $\quad a_2 = 2 \cdot \frac{1}{\sqrt{3}} - 1 \cdot \frac{1}{\sqrt{3}} + 3 \cdot \frac{1}{\sqrt{3}} = \frac{4}{\sqrt{3}}$,

$a_3 = 2 \cdot \left(-\frac{1}{\sqrt{6}}\right) + 1 \cdot \frac{1}{\sqrt{6}} + 3 \cdot \frac{2}{\sqrt{6}} = \frac{5}{\sqrt{6}}$. $\quad$ Hence $x = \frac{3}{\sqrt{2}} v_1 + \frac{4}{\sqrt{3}} v_2 + \frac{5}{\sqrt{6}} v_3$.

- So it is useful to have an orthonormal basis.
  But we still need to show that it always exists! (in a fin. dim. inner product space)

Ex Let's consider a simple case first.
- Suppose $\{w_1, w_2\}$ is a lin. indep subset of an inner product space (and hence a basis for $W = \mathrm{Span}(\{w_1, w_2\})$).
- We want to construct an orthogonal set from $\{w_1, w_2\}$ that spans the same subspace W.



The picture above suggests that the set $\{v_1, v_2\}$ with $v_1 = w_1$, $v_2 = w_2 - cw_1$ has this property if c is chosen so that v_2 is orthogonal to w_1.
To find c, we need to solve the equation
$0 = \langle v_2, w_1 \rangle = \langle w_2 - cw_1, w_1 \rangle = \langle w_2, w_1 \rangle - c \langle w_1, w_1 \rangle$.
So $c = \dfrac{\langle w_2, w_1 \rangle}{\|w_1\|^2}$, and $v_2 = w_2 - \dfrac{\langle w_2, w_1 \rangle}{\|w_1\|^2} w_1$

The next theorem show that this process can be extended to any finite lin. indep. subset.

Thm 6.4 Let V be an inner prod. space and $S = \{w_1, \dots, w_n\}$ a lin. indep subset of V.
Define $S' = \{v_1, \dots, v_n\}$, where $v_1 = w_1$ and
$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j \quad \text{for } 2 \le k \le n. \qquad (1)$$
Then S' is an orthogonal set of non-zero vectors such that $\mathrm{Span}(S') = \mathrm{Span}(S)$.

Proof. By induction on n, the number of vectors in S.
For $k = 1, 2, \dots, n$, let $S_k = \{w_1, \dots, w_k\}$.
If $n=1$, then the theorem is proved by taking $S_1' = S_1$, i.e. $v_1 = w_1 \neq 0$.
For $n>1$. Assume that the set
$S'_{k-1} = \{v_1, \dots, v_{k-1}\}$ with the desired properties has been constructed by the repeated use of (1).

We show that $S'_k = \{v_1, \dots, v_{k-1}, v_k\}$ also has the desired properties, where v_k is obtained from S'_{k-1} by (1).

If $v_k = 0$, then (1) implies that $w_k \in \mathrm{Span}(S'_{k-1}) \overset{\text{inductive hyp.}}{=} \mathrm{Span}(S_{k-1})$, which contradicts the assumption that S_k is lin. indep. Hence $v_k \neq 0$.

For $1 \le i \le k-1$ it follows from (1) that

$$\langle v_k, v_i\rangle = \langle w_k, v_i\rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} \underbrace{\langle v_j, v_i\rangle}_{\substack{=0 \text{ for } i\neq j \text{ by ind. hyp. on} \\ \text{orthogonality of } S'_{k-1}.}} = \langle w_k, v_i\rangle - \frac{\langle w_k, v_i\rangle}{\|v_i\|^2}\|v_i\|^2 = 0$$

Hence S'_k is an orthogonal set of non-zero vectors.

By (1), $\mathrm{Span}(S'_k) \subseteq \mathrm{Span}(S_k)$. By Cor. 2 to Thm 6.3, S'_k is lin. indep.

So $\dim(\mathrm{Span}(S'_k)) = \dim(\mathrm{Span}(S_k)) = k$. Therefore $\mathrm{Span}(S'_k) = \mathrm{Span}(S_k)$.

The construction of $\{v_1, \dots, v_n\}$ by the use of Thm 6.4 is called the Gram-Schmidt process.

Ex Let $V = \mathbb{R}^4$ (with the standard inner prod.), $w_1 = (1,0,1,0)$, $w_2 = (1,1,1,1)$, $w_3 = (0,1,2,1)$. Then $\{w_1, w_2, w_3\}$ is lin. indep.

We use the G-S process to compute the orthogonal vectors v_1, v_2, v_3.

Take $v_1 = w_1 = (1,0,1,0)$. Then

$$v_2 = w_2 - \frac{\langle w_2, v_1\rangle}{\|v_1\|^2} v_1 = (1,1,1,1) - \frac{(1\cdot 1 + 0\cdot 1 + 1\cdot 1 + 0\cdot 1)}{\left(\sqrt{1^2+0^2+1^2+0^2}\right)^2}(1,0,1,0) = (0,1,0,1).$$

$$v_3 = w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} v_2 = (0,1,2,1) - \frac{2}{2}(1,0,1,0) - \frac{2}{2}(0,1,0,1) = (-1,0,1,0).$$

Now we normalize them to obtain the orthonormal basis $\{u_1, u_2, u_3\}$ where

$$u_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{\sqrt{2}}(1,0,1,0)$$

$$u_2 = \frac{1}{\|v_2\|} v_2 = \frac{1}{\sqrt{2}}(0,1,0,1)$$

$$u_3 = \frac{v_3}{\|v_3\|} = \frac{1}{\sqrt{2}}(-1,0,1,0).$$

Thm 6.5 Let V be a non-zero inner prod. space, $\dim(V) < \infty$.
Then V has an orthonormal basis β.
Furthermore, if $\beta = \{v_1, \dots, v_n\}$ and $x \in V$, then $x = \sum_{i=1}^{n} \langle x, v_i\rangle v_i$.

Proof.
Let β_0 be an ordered basis for V.
Applying Thm 6.4, we obtain an orthogonal set β' of non-zero vectors with $\mathrm{Span}(\beta_0) = \mathrm{Span}(\beta') = V$.
Normalizing each vector in β', we obtain an orthonormal set β with $\mathrm{Span}(\beta) = \mathrm{Span}(\beta') = V$.
By Corollary 2 (to Thm 6.3), β is lin. indep — hence an orthonormal basis for V. The rest follows by Corollary 1.

Cor Let V be an inner prod. space with an orthonormal basis $\beta = \{v_1, \dots, v_n\}$.
Let T be a lin. operator on V, and let $A = [T]_\beta$.
Then for any i, j, $A_{ij} = \langle T(v_j), v_i\rangle$.

Orthogonal complement

Def Let $S \subseteq V$ be non-empty, V – inner prod. space.
Let $S^\perp$ ("S perp") be the set of all vectors in V that are orthogonal to every vector in S. That is,

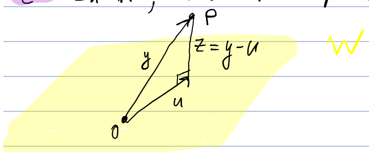
$$S^\perp = \{x \in V : \langle x, y\rangle = 0 \text{ for all } y \in S\}.$$ – the orthogonal complement of S.

Note: $S^\perp$ is a subspace of V for any $S \subseteq V$.

Ex

1) $\{0\}^\perp = V$ and $V^\perp = \{0\}$ for any inner prod. space
2) If $V = \mathbb{R}^3$ and $S = \{e_3\}$, then $S^\perp$ equals the xy-plane.

Ex In $\mathbb{R}^3$, consider a point P and a plane W. How to find the distance from P to W?



By the picture, can be restated as:
Determine the vector $u \in W$ that is "closest" to y, the distance given by $\|y-u\|$. Notice: $z = y-u$ is orthogonal to every vector in W, so $z \in W^\perp$.

We can find u as follows.

Thm 6.6 Let W be a subspace of an inner prod. space V, $\dim(W) < \infty$. Let $y \in V$. Then there exists unique vectors $u \in W$ and $z \in W^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, \dots, v_k\}$ is an orthonormal basis for W, then $u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$.

Proof. Let $\{v_1, \dots, v_k\}$ and u be as above.
Let $z = y - u$. Then $u \in W$ and $y = u + z$. (see HW).
To show that $z \in W^\perp$, it suffices to show that z is orthogonal to each v_j.
For any j we have:

$$\langle z, v_j \rangle = \left\langle \left(y - \sum_{i=1}^{k} \langle y, v_i \rangle \right), v_j \right\rangle = \langle y, v_j \rangle - \sum_{i=1}^{k} \langle y, v_i \rangle \langle v_i, v_j \rangle =$$

$$= \langle y, v_j \rangle - \langle y, v_j \rangle = 0.$$

To show uniqueness of u and z, suppose that $y = u + z = u' + z'$, where $u' \in W$, $z' \in W^\perp$.
Then $u - u' = z' - z \in W \cap W^\perp = \{0\}$. Therefore, $u = u'$ and $z = z'$.

Corollary In the notation of Thm 6.6, the vector u is the unique vector in W that is "closest" to y. That is, for any $x \in W$, $\|y - x\| \geq \|y - u\|$, and the equality holds if and only if $x = u$.

Proof
See Textbook, p. 350.

This vector u in the corollary is called the orthogonal projection of y on W.

Thm 6.7 Let $S = \{v_1, \dots, v_k\} \subseteq V$ be orthonormal, $\dim(V) = n$. Then

a) S can be extended to an orthonormal basis $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ for V.
b) If $W = \mathrm{Span}(S)$, then $S_1 = \{v_{k+1}, \dots, v_n\}$ is an orthonormal basis for $W^\perp$.
c) If W is any subspace of V, then $\dim(V) = \dim(W) + \dim(W^\perp)$.

Proof

a) By Cor. 2 to the replacement thm, S can be extended to an ordered basis $S' = \{v_1, \dots, v_k, w_{k+1}, \dots, w_n\}$ for V. Now apply the Gram-Schmidt process to S'.
The first k vectors resulting from this process are the vectors in S, and this new set spans V.
Normalize the last $n-k$ vectors.

b) Because S_1 is a subset of a basis, it is lin. indep.
Since S_1 is clearly a subset of $W^\perp$, we need only show that $\mathrm{Span}(S_1) = W^\perp$.
For any $x \in V$,
$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$.

If $x \in W^\perp$, then $\langle x, v_i \rangle = 0$ for $1 \le i \le k$. Therefore,

$$x = \sum_{i=k+1}^{n} \langle x, v_i \rangle v_i \in \mathrm{Span}(S_1).$$

c) Let W be a subspace of V. It is a fin. dim. inner prod. space because V is, so has an orthonorm. basis $\{v_1, \dots, v_k\}$. By (a) and (b),

$$\dim(V) = n = k + (n-k) = \dim(W) + \dim(W^\perp).$$

Ex.

Let $W = \mathrm{Span}(\{e_1, e_2\})$ in F^3. Then $x = (a,b,c) \in W^\perp \iff 0 = \langle x, e_1 \rangle = a$ and $0 = \langle x, e_2 \rangle = b$.
So $x = (0,0,c)$, and $W^\perp = \mathrm{Span}(\{e_3\})$.

The adjoint of a lin. operator

If V is an inner prod. space, then for any $y \in V$ the function $g: V \to F$ defined by $g(x) = \langle x, y \rangle$ is linear. If V is finite dimensional, then every lin. transformation from V to F is of this form:

Thm 6.8 Let V be a fin. dim. inn. prod. space over F, and let $g: V \to F$ be a lin. transf. Then there exists a unique vector $y \in V$ s.t. $g(x) = \langle x, y \rangle$ for all $x \in V$.

Proof

If $\beta = \{v_1, \dots, v_n\}$ an orthonorm. basis for V, can take $y = \sum_{i=1}^{n} \overline{g(v_i)}\, v_i$. (see Textbook, p. 357 for a proof.)

Ex

Define $g: \mathbb{R}^2 \to \mathbb{R}$ by $g(a_1, a_2) = 2a_1 + a_2$, g is lin.

Let $\beta = \{e_1, e_2\}$ – an orthonormal basis, and let $y = g(e_1)e_1 + g(e_2)e_2 = 2e_1 + e_2 = (2,1)$.

Then $g(a_1, a_2) = \langle (a_1, a_2), (2,1) \rangle = 2a_1 + a_2$.

Thm 6.9 Let V be a fin. dim. inn. prod. space and $T \in \mathcal{L}(V)$. Then there exists a unique lin. operator $T^* \in \mathcal{L}(V)$ s.t.

$\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ for all $x, y \in V$.

(See Textbook, p. 358 for a proof.)

T^* is called the adjoint of T.

Remark 1) We also have $\langle x, T(y) \rangle = \overline{\langle T(y), x \rangle} = \overline{\langle y, T^*(x) \rangle} = \langle T^*(x), y \rangle$ for all $x, y \in V$.

2) If V is inf. dim, then the existence of the adjoint is not guaranteed (See Ex. 24, Sec. 6.4).

Recall that for a matrix $A \in M_{m \times n}(F)$, the adjoint of A is defined as the matrix $A^* \in M_{n \times m}(F)$ s.t. $A^*_{ij} = \overline{A_{ji}}$ for all i, j. So if $F = \mathbb{R}$, then simply $A^* = A^t$.

Thm 6.10 Let V be a fin. dim. inn. prod. space and β an orthonorm. basis for V. If $T \in \mathcal{L}(V)$, then

$[T^*]_\beta = [T]_\beta^*$.

Proof

Let $A = [T]_\beta$, $B = [T^*]_\beta$ and $\beta = \{v_1, \dots, v_n\}$. From the corollary to Thm 6.5 we have:

$$B_{ij} = \langle T^*(v_j), v_i \rangle \underset{\substack{\text{"symmetry" of} \\ \text{inner product}}}{=} \overline{\langle v_i, T^*(v_j) \rangle} \underset{\substack{\text{def. of} \\ \text{adjoint}}}{=} \overline{\langle T(v_i), v_j \rangle} = \overline{A_{ji}} = (A^*)_{ij}.$$ Hence $B = A^*$.

Cor Let $A \in M_{n \times n}(F)$. Then $L_{A^*} = (L_A)^*$.

Proof.

If β is the standard ordered basis for F^n, then $[L_A]_\beta = A$. Hence

$[(L_A)^*]_\beta = [L_A]_\beta^* = A^* = [L_{A^*}]_\beta$, so $(L_A)^* = L_{A^*}$.

Ex

Let $T \in \mathcal{L}(\mathbb{C}^2)$ be defined by $T(a_1, a_2) = (2ia_1 + 3a_2, a_1 - a_2)$.

If β is the standard ordered basis for $V = \mathbb{C}^2$, $\beta = \{e_1, e_2\}$, $e_1 = \begin{pmatrix}1\\0\end{pmatrix}$, $e_2 = \begin{pmatrix}0\\1\end{pmatrix}$, then

$[T]_\beta = \begin{pmatrix} | & | \\ [T(e_1)]_\beta & [T(e_2)]_\beta \\ | & | \end{pmatrix} = \begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix}$. So $[T^*]_\beta = [T]_\beta^* = \begin{pmatrix} -2i & 1 \\ 3 & -1 \end{pmatrix}$.

Hence $T^*(a_1, a_2) = [T]_\beta^* \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = (-2ia_1 + a_2, 3a_1 - a_2)$.

There are many algebraic analogies between the conjugates of complex numbers and the adjoints of lin. operators, e.g. $(T+U)^* = T^* + U^*$, $(cT)^* = \bar{c}T^*$, etc. (see Thm 6.11 in the book).

Normal and self-adjoint operators

In a vector space V:

$T \in \mathcal{L}(V)$ is diag'z $\iff$ $[T]_\beta$ is diagonal for some basis β for V $\iff$ V has a basis of e.vects for T

In an inn. prod. space V:

??? $\iff$ V has an orthonormal basis of e.vects for T.

If such an orthonorm. basis β exists, then $[T]_\beta$ is a diagonal matrix. Hence $[T^*]_\beta = [T]_\beta^*$ is also diagonal. Because diagonal matrices commute, T and T^* commute. This motivates:

Def 1) Let V be an inn. prod. space, let $T \in \mathcal{L}(V)$.
T is normal if $TT^* = T^*T$.
2) $A \in M_{n \times n}(F)$, for $F = \mathbb{R}, \mathbb{C}$, is normal if $AA^* = A^*A$.

By Thm 6.10, T is normal $\iff$ $[T]_\beta$ is normal.

Thm 6.16 Let $T \in \mathcal{L}(V)$ for V a fin. dim. complex inn. prod. space (so $F = \mathbb{C}$)
Then T is normal $\iff$ exists an orthonorm. basis for V of e.vects. for T.

This solves our problem when $F = \mathbb{C}$, but not for $F = \mathbb{R}$:

Ex Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by θ, $0 < \theta < \pi$. If β - stand. ord. basis,

$$[T]_\beta = A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that $AA^* = I = A^*A$. So A (and T) are normal.

However, T has no e.vects at all! (exercise).

For real inn. prod. spaces, need a stronger condition.

Def 1) $T \in \mathcal{L}(V)$ is self-adjoint (Hermitian) if $T = T^*$.
2) $A \in M_{n \times n}(F)$ is self-adjoint (Hermitian) if $A = A^*$.

Lemma Let $T \in \mathcal{L}(V)$ be self-adjoint. Then:
1) Every e.val. of T is real. (even when $F = \mathbb{C}$!)
2) If $F = \mathbb{R}$, then the char. poly. of T splits.

Thm 6.17
Let $T \in \mathcal{L}(V)$, V - fin. dim. real inn. prod. space. Then:
T is self-adjoint $\iff$ exists an orthonorm. basis β for V of e.vects. for T.