Pseudo-Definition. A **field** is an "algebraic system" $\mathbb{F}$ having:

(1) elements 0,1 (and possibly others)
(2) operations $+, \times, -$, and $^{-1}$ (the last defined for all nonzero elements)

and satisfying the "obvious" algebraic laws. (See Appendix C for the real definition.)

Example 1. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Z}_p$ (p a prime) are fields.

Let $\mathbb{F}$ be a field.

Definition. A **vector space over** $\mathbb{F}$ is $\boxed{\text{a set}}$ V on which $\boxed{\text{two operations}}$

- **addition**, $V \times V \to V$, denoted $x + y$
- **scalar multiplication**, $\mathbb{F} \times V \to V$, denoted ax

$\boxed{\text{are defined}}$, and such that $\boxed{\text{the following conditions hold}}$:

For all $x, y, z \in V$ and $a, b \in \mathbb{F}$:

(VS 1) $x + y = y + x$
(VS 2) $(x + y) + z = x + (y + z)$
(VS 3) There exists a "zero vector" in V, denoted 0, which satisfies $x + 0 = x$ for all $x \in V$.
(VS 4) For every $x \in V$ there exists $u \in V$ satisfying $x + u = 0$.
(VS 5) $1x = x$
(VS 6) $(ab)x = a(bx)$
(VS 7) $a(x + y) = ax + ay$
(VS 8) $(a + b)x = ax + bx$

To **define** a vector space, you must specify the set and the two operations.
To **prove** that a set with two operations is a vector space, you need to verify the 8 conditions.

Example 2. $\mathbb{R}^n$ is the set of all n-tuples $(a_1, a_2, \ldots, a_n)$ of real numbers. Addition and scalar multiplication (by real numbers) on $\mathbb{R}^n$ are defined "coordinate-wise," i.e.,

$$\begin{aligned} (a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) &\overset{def}{=} (a_1{+}b_1, a_2{+}b_2, \ldots, a_n{+}b_n) \\ c(a_1, a_2, \ldots, a_n) &\overset{def}{=} (ca_1, ca_2, \ldots, ca_n). \end{aligned}$$

Claim. *$\mathbb{R}^n$ with coordinate-wise addition and scalar multiplication is a vector space over $\mathbb{R}$.*

Proof sketch. (Omitted) □

Example 3. More generally, for any field $\mathbb{F}$, the set $\mathbb{F}^n = \{(a_1, a_2, \ldots, a_n) : a_1, \ldots, a_n \in \mathbb{F}\}$ of all n-tuples from $\mathbb{F}$, with coordinatewise addition and scalar multiplication, is a vector space over $\mathbb{F}$.

Example 4. For any nonempty set D, $\mathbb{F}^D$ is the set of all functions $D \to \mathbb{F}$. Given two functions $f, g \in \mathbb{F}^D$ and $a \in \mathbb{F}$, define the functions $f + g$ and af "pointwise" by

$$\begin{aligned} (f + g)(x) &\overset{def}{=} f(x) + g(x), \\ (af)(x) &\overset{def}{=} a \cdot f(x), \qquad x \in D. \end{aligned}$$

Claim. *For any nonempty set D, $\mathbb{F}^D$ with pointwise operations is a vector space over $\mathbb{F}$.*

Example 5. Let $V = \{⚽\}$ where ⚽ is the apple I brought to class. Defining addition and scalar multiplication in the only possible way, V is a vector space over $\mathbb{F}$ (for any field $\mathbb{F}$).

More examples. Let $\mathbb{F}$ be a field.

(1) For $n \geq 0$, $\mathsf{P}_n(\mathbb{F})$ denotes the set of all "formal polynomials" in the variable x, of degree at most n, using coefficients from $\mathbb{F}$. Thus

$$\mathsf{P}_n(\mathbb{F}) = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \,:\, a_0, a_1, \ldots, a_n \in \mathbb{F}\}.$$

Addition of polynomials in $\mathsf{P}_n(\mathbb{F})$ is defined "term-wise" in the usual way (using the arithmetic of $\mathbb{F}$). Multiplication of a polynomial by a scalar $c \in \mathbb{F}$ is defined similarly.

Claim. *$\mathsf{P}_n(\mathbb{F})$ with addition and scalar multiplication defined "term-wise" is a vector space over $\mathbb{F}$.*

(2) $\mathbb{F}[x]$ is the set of <u>all</u> polynomails in x with coefficients from $\mathbb{F}$. Addition and scalar multiplication is the same is in the previous example.

Claim. *$\mathbb{F}[x]$ is a vector space over $\mathbb{F}$.*

Remark. The text uses the notation $\mathsf{P}(\mathbb{F})$ instead of $\mathbb{F}[x]$, but this is nonstandard.

Next: some basic facts true of all vector spaces.

Theorem (Cancellation Law)**.** *Suppose V is a vector space. If $x, y, z \in V$ and $x + z = y + z$, then $x = y$.*

Proof. (Omitted) □

Corollary 1. *Suppose V is a vector space. There is exactly one vector in V that can be the zero vector.*

Proof. (Omitted) □

Corollary 2. *Suppose V is a vector space and $x \in V$. There is exactly one vector $u \in V$ satisfying $x + u = 0$.*

Proof. Just like the proof of Corollary 1. □

Definition. Let V be a vector space and $x, y \in V$.

(1) $-x$ denotes the unique vector $u \in V$ satisfying $x + u = 0$.
(2) $x - y$ denotes $x + (-y)$.

Because of (VS 2), we can (and do) write expressions like $x_1 + x_2 + \cdots + x_n$ without declaring where the brackets go. And we could (if required) prove true facts like

$$c(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \;=\; (ca_1)x_1 + (ca_2)x_2 + \cdots + (ca_n)x_n.$$

Definition. Let V be a vector space over $\mathbb{F}$ and suppose $x, u_1, \ldots, u_n \in V$. We say that x is a **linear combination of** $u_1, \ldots, u_n$ if there exist scalars $a_1, \ldots, a_n \in \mathbb{F}$ satisfying

$$x = a_1u_1 + a_2u_2 + \cdots + a_nu_n.$$

Basic Problem: Given $x, u_1, \ldots, u_n \in V$, to determine whether x is a linear combination of $u_1, \ldots, u_n$.

Example 3. Consider the vector space $\mathbb{R}[x]$ of formal polynomials over $\mathbb{R}$. Is $4x^4 + 7x^2 - 2x + 3$ a linear combination of

$$x^4 - x^2 + 2x - 1, \qquad 2x^4 + 3x^2 + 2x, \qquad x^4 + 4x^2 + 1, \qquad 2x^2 + 3, \qquad x^4 + 1 \text{ ?}$$

Theorem. *Suppose V is a vector space over $\mathbb{F}$, $x \in V$, and $a \in \mathbb{F}$.*

(1) $0x = 0$.
(2) $(-a)x = -(ax) = a(-x)$.
(3) $a0 = 0$.

Remark. Note the overloaded notation in the statement.

Definition. Let V be a vector space over $\mathbb{F}$, and let $S \subseteq V$.

(1) V is <u>closed under addition</u> if $x, y \in S \implies x + y \in S$.
(2) V is <u>closed under scalar multiplication</u> if $x \in S$ and $a \in \mathbb{F} \implies ax \in S$.

Definition. Let V be a vector space over $\mathbb{F}$. A subset W of V is a called **subspace** of V if

(1) W is closed under the operations of V, and
(2) $W \neq \varnothing$.

Theorem. *Suppose V is a vector space over $\mathbb{F}$ and W is a subspace of V. Then W, <u>with the operations of V restricted to W</u>, is a vector space over $\mathbb{F}$.*

Proof sketch. (VS 1), (VS 2), and (VS 5)–(VS 8) follow automatically because of their logical nature (universally quantified statements). Proving (VS 3) and (VS 4) requires a little more work and can be done using the previous Theorem; in particular, $-x = -(1x) = (-1)x$. $\square$

Remark. The converse to the previous theorem is also true: if $W \subseteq V$ and W with the operations of V restricted to W is a vector space, then $W \neq \varnothing$ and W is closed under the operations of V (so is a subspace of V).

Definition. Suppose V is a vector space over $\mathbb{F}$, $x \in V$, and $\varnothing \neq S \subseteq V$.

(1) x is a **linear combination of** S if x is a linear combination of some finite list of vectors from S. (Note that S might be infinite.)

(2) The **span of** S, written $\operatorname{span}(S)$, is the set of all vectors $x \in V$ which are linear combinations of S.

We also define $\operatorname{span}(\varnothing) = \{0\}$.

Example 1. In $\mathbb{R}[x]$, what is the span of the infinite set $S = \{x, x^2, x^3, x^4, \ldots\}$? It includes all linear combinations of finitely many of $x, x^2, x^3, x^4, \ldots$. Thus we get all polynomials of the form

$$a_1 x + a_2 x^2 + \cdots + a_n x^n, \qquad a_1, a_2, \ldots, a_n \in \mathbb{R}.$$

In other words, $\operatorname{span}(S) = \{f(x) \in \mathbb{R}[x] \,:\, f(0) = 0\}$.

Technical Observations. (Assume $S \neq \varnothing$.)

(1) Suppose $x \in \operatorname{span}(S)$. So x is a linear combination of some finite list $u_1, \ldots, u_m$ from S, say,

$$x = a_1 u_1 + \cdots + a_m u_m.$$

If $v_1, \ldots, v_n$ are some more vectors from S, then x is also a linear combination of $u_1, \ldots, u_m, v_1, \ldots, v_n$, since we can write

$$x = a_1 u_1 + \cdots + a_m u_m + 0 v_1 + \cdots + 0 v_n.$$

(2) Thus if S is finite, say $S = \{u_1, \ldots, u_n\}$, then $x \in \operatorname{span}(S)$ iff x is a linear combination of $u_1, \ldots, u_n$.

(3) If S is infinite, we can say the following. Suppose $x, y \in \operatorname{span}(S)$. Then x is a linear combination of a finite list $u_1, \ldots, u_m$ from S and y is a linear combination of a finite list $v_1, \ldots, v_n$ from S. By the earlier remark, we can view both x and y as linear combinations of the **same** list $u_1, \ldots, u_m, v_1, \ldots, v_n$.

Theorem (On span). *Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$. Then* $\operatorname{span}(S)$ *is the (unique) smallest subspace of V which contains S. That is,*

(1) $\operatorname{span}(S)$ *is a subspace of V.*

(2) $\operatorname{span}(S) \supseteq S$.

(3) *If W is any subspace of V and $W \supseteq S$, then $W \supseteq \operatorname{span}(S)$.*

Proof. (1) and (2) in class; (3) deferred to Wednesday's lecture. $\square$

Today: redundancies in span.

Example 1. Suppose $S = \{u_1, u_2, u_3, u_4, u_5\}$ and u_3 can be written as a linear combination of u_2, u_4, u_5, say

$$u_3 = c_2u_2 + c_4u_4 + c_5u_5.$$

Claim: $\operatorname{span}(S) = \operatorname{span}(S \setminus \{u_3\})$.

Proof. $\supseteq$ can be quickly proved using the Theorem on span from Monday's lecture. For $\subseteq$, argue directly. $\square$

Also note that

$$0u_1 + a_2u_2 + (-1)u_3 + a_4u_4 + a_5u_5 = 0.$$

The scalars $0, a_2, -1, a_4, a_5$ are not all 0 (because of -1). This motivates the formal definition.

Definition. Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$. We say that S is **linearly dependent** if there exist distinct vectors $u_1, \ldots, u_n \in S$ and scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that

(1) $a_1u_1 + \cdots + a_nu_n = 0$, and
(2) $a_1, \ldots, a_n$ are not all 0.

If S is not linearly dependent, we say it is **linearly independent**.

Let's explore this. A set is S linearly dependent

$$\iff \quad (\exists \text{ distinct } u_1, \ldots, u_n \in S)(\exists a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n = 0 \text{ and } \neg(a_1 = \cdots = a_n = 0))$$

Thus S is linearly independent

$$\begin{aligned}
&\iff \neg(\exists \text{ distinct } u_1, \ldots, u_n \in S)(\exists a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n = 0 \text{ and } \neg(a_1 = \cdots = a_n = 0)) \\
&\iff (\forall \text{ distinct } u_1, \ldots, u_n \in S)(\forall a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n \neq 0 \text{ or } a_1 = \cdots = a_n = 0) \\
&\iff (\forall \text{ distinct } u_1, \ldots, u_n \in S)(\forall a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n = 0 \implies a_1 = \cdots = a_n = 0)
\end{aligned}$$

Technical Observation. Suppose S is finite and nonempty, say $S = \{u_1 \ldots, u_n\}$. Then the definition of linear dependence, and the characterization of linear independence, can both be simplified by dropping the "$\exists$ distinct $u_1, \ldots, u_n \in S$" or "$\forall$ distinct $u_1, \ldots, u_n \in S$." Thus (in this situation),

- S is linearly dependent iff
$$(\exists a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n = 0 \text{ and } \neg(a_1 = \cdots = a_n = 0)).$$
- S is linearly independent iff
$$(\forall a_1, \ldots, a_n \in \mathbb{F})(a_1u_1 + \cdots + a_nu_n = 0 \implies a_1 = \cdots = a_n = 0).$$

Question: Is $S = \varnothing$ linearly dependent, or linearly independent? $\boxed{\text{Linearly independent}}$

Question: Is $S = \{0\}$ linearly dependent, or linearly independent? $\boxed{\text{Linearly dependent}}$

Theorem (On dependence)**.** *Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$. S is linearly dependent iff $S = \{0\}$ or some vector in S is a linear combination of other vectors in S.*

Recall: if V is a vector space/$\mathbb{F}$ and $S \subseteq V$, then:

(1) $\operatorname{span}(S)$ is the set of all linear combinations of S.
(2) S is linearly dependent if there exist distinct $u_1, \dots, u_n \in S$ and there exist $a_1, \dots, a_n \in \mathbb{F}$, not all zero, such that $a_1u_1 + \cdots + a_nu_n = 0$.
   - Else S is linearly independent.

Definition. Let V be a vector space over $\mathbb{F}$.

(1) A subset $S \subseteq V$ is a spanning set if $\operatorname{span}(S) = V$. We also say S spans V.
(2) We say V is finitely [countably] spanned if V has a finite [countable] spanning set.

Here *countable* means "finite or in 1-1 correspondence with $\mathbb{N}$."

E.g., $\mathbb{F}^n$ is finitely spanned, e.g., by $\{e_1, \dots, e_n\}$ where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ (1 in ith spot).
$\mathbb{F}[x]$ is countably spanned, e.g., by $\{1, x, x^2, \dots, x^n, \dots\}$.
$\mathbb{R}^{[0,1]}$ is not countably spanned.

Definition. Let V be a vector space. A **basis** for V is a subset $S \subseteq V$ which is linearly independent and spans V.

Theorem. *Every countably spanned vector space has a basis.*

Proof. Let V be spanned by the countable set S; so $S = \{v_1, \dots, v_n\}$ or $S = \{v_1, v_2, \dots\}$.
We can assume WLOG that $0 \notin S$. Define

$$T = \{v_j \,:\, v_j \notin \operatorname{span}(\{v_1, \dots, v_{j-1}\})\}.$$

Write $T = \{v_{i_1}, v_{i_2}, \dots\}$ (finite or infinite), $i_1 < i_2 < \cdots$.

Claim: T is a basis for V.

First show T is linearly independent. Argue by contradiction; assume T is linearly dependent. We can choose some finite initial segment of it, say $\{v_{i_1}, \dots, v_{i_k}\}$ for which we can choose $a_1, \dots, a_k \in \mathbb{F}$, not all 0, with

$$a_1v_{i_1} + \cdots + a_kv_{i_k} = 0.$$

Assume k has been chosen to be smallest with this property.
Cannot have $k = 1$ (since $0 \notin S$). So $k > 1$.
If $a_k = 0$, then

$$a_1v_{i_1} + \cdots + a_{k-1}v_{i_k-1} = 0, \qquad \text{not all } a_1, \dots, a_{k-1} \text{ equal } 0$$

Contradicts choice of k; proves $a_k \neq 0$.

Manipulate to get v_{i_k} a linear combo of $v_{i_1}, \dots, v_{i_{k-1}}$; contradicts $v_{i_k} \in T$.
Next we must show $\operatorname{span}(T) = V$. Intuition: $\operatorname{span}(S) = V$ and we've only thrown out "redundant" vectors to get T. Formally, list vectors of S and of T. For each n, let

$$\begin{aligned} S_n &= \{v_1, \dots, v_n\} \\ T_n &= \{v_{i_k} \in T \,:\, i_k \leq n\}. \end{aligned}$$

Argue, by induction on n, that $\operatorname{span}(T_n) = \operatorname{span}(S_n)$ for all n. (Don't do.)
Then to show $\operatorname{span}(T) = V$, equivalently, $V \subseteq \operatorname{span}(T)$, let $x \in V$; then $x \in \operatorname{span}(S)$; so by picking n large enough we get $x \in \operatorname{span}(S_n) = \operatorname{span}(T_n) \subseteq \operatorname{span}(T)$. $\square$

Jan 17, Lecture 7: Dimensions

The next main result is to prove that any two bases of a **finitely spanned** vector space have the same number of elements. To prove this, we first have an intermediate result.

Theorem 1.10. *Let V be a finitely spanned vector space over a field $\mathbb{F}$. Let $\{v_1, \ldots, v_m\}$ be a basis for V. Let $\{w_1, \ldots, w_n\} \subset V$ and $n > m$. Then $\{w_1, \ldots, w_n\}$ is a linearly dependent set.*

Proof Sketch. Idea: Replace successively $v_1, v_2, \ldots, v_r$ so that $w_1, w_2, \ldots, w_r, v_{r+1}, \ldots, v_m$ generate V for all $1 \le r \le m-1$. Finally, we have $w_1, \ldots, w_m$ generate V. Below is the detail of the proof.

Assume that $\{w_1, \ldots, w_n\}$ is linearly independent. (*)

Statement: After renumbering $v_1, \ldots, v_m$ if necessary, we have $w_1, w_2, \ldots, w_r, v_{r+1}, \ldots, v_m$ generate V for all $1 \le r \le m-1$.

- Base step: Since $\{v_1, \ldots, v_m\}$ is a basis, we have

$$w_1 = a_1 v_1 + \cdots + a_m v_m$$

  By assumption, $w_1 \neq 0$, so $a_i \neq 0$ for some $1 \le i \le m$. After renumbering $v_1, \ldots, v_m$ if necessary, we may assume WLOG that $a_1 \neq 0$. Then we can solve for v_1 and get

$$a_1 v_1 = w_1 - a_2 v_2 - \cdots - a_m v_m$$

$$v_1 = a_1^{-1} w_1 - a_1^{-1} a_2 v_2 - \cdots - a_1^{-1} a_m v_m$$

$$V = \operatorname{span}(\{v_1, v_2, \ldots, v_m\}) \subset \operatorname{span}(\{w_1, v_2, \ldots, v_m\}) \subset V$$

$$V = \operatorname{span}(\{w_1, v_2, \ldots, v_m\}).$$

- Assume by induction that there is an integer r with $1 \le r \le m-1$ such that, after a suitable renumbering of $v_1, \ldots, v_m$, $V = \operatorname{span}(\{w_1, \ldots, w_r, v_{r+1}, \ldots, v_m\})$.

- We will prove the statement is true for $r+1$, that is, $V = \operatorname{span}(\{w_1, \ldots, w_r, w_{r+1}, v_{r+2}, \ldots, v_m\})$, after a suitable renumbering of $v_1, \ldots, v_m$.
  Since $w_1, \ldots, w_r, v_{r+1}, \ldots, v_m$ generate V, we have

$$w_{r+1} = b_1 w_1 + \cdots + b_r w_r + c_{r+1} v_{r+1} + \cdots + c_m v_m.$$

  - Claim: We cannot have $c_j = 0$ for all $j = r+1, \ldots m$. Indeed, if $c_{r+1} = \cdots = c_m = 0$, then w_{r+1} is a linear combination of $w_1, \ldots, w_r$, hence $\{w_1, \ldots, w_{r+1}, \ldots, w_n\}$ is linearly dependent, a contradiction with the assumption (*).

- WLOG, assume $c_{r+1} \neq 0$. Then

$$v_{r+1} = c_{r+1}^{-1} w_{r+1} - c_{r+1}^{-1} b_1 w_1 - \cdots - c_{r+1}^{-1} b_r w_r - c_{r+1}^{-1} c_{r+2} v_{r+2} - \cdots - c_{r+1}^{-1} c_m v_m$$

Using the same argument as in the base step and using the induction assumption, we have

$$V = \operatorname{span}(\{w_1, \ldots, w_r, w_{r+1}, v_{r+2}, \ldots, v_m\}).$$

- So by induction, we have proved that $w_1, \ldots, w_m$ generates V. If $n > m$, then we can write w_n as a linear combination of $w_1, \ldots, w_m$. Therefore, the set $\{w_1, \ldots, w_n\}$ is linearly dependent, a contradiction to the assumption (*). In conclusion, the assumption (*) is wrong, and the set $\{w_1, \ldots, w_n\}$ is linearly dependent.

□

Theorem 1.11. *Let V be a vector space and suppose that one basis has n elements, and another basis has m elements. Then $m = n$.*

Proof. Previous theorem implies that both alternatives $n > m$ and $m > n$ are impossible, and hence $m = n$. □

Definition 15. *Let V be a vector space having a basis consisting of n elements. We say that n is the dimension of V, $\dim V = n$.*
For $V = \{0\}$, $\dim\{0\} = 0$.
A vector space which has a basis consisting of a finite number of elements , or the zero vector space, is called finite dimensional. Other vector space are called infinite dimensional.

Example 14. Dimensions of $\mathbb{F}^n, M_{m\times n}(\mathbb{F}), P_n(\mathbb{F})$.

Let $\{v_1, \ldots, v_n\}$ be linearly independent elements of a vector space V. We say that $\{v_1, \ldots, v_n\}$ is a maximal set of linearly independent elements of V if given any $w \in V$, the set $\{w, v_1, \ldots, v_m\}$ is linearly dependent.
Corollary: Let V be a vector space.

- If $\{v_1, \ldots, v_n\}$ is a maximal set of linearly independent elements of V, then $\{v_1, \ldots, v_n\}$ is a basis of V.

- If $\dim V = n$ and $\{v_1, \ldots, v_n\}$ is a linearly independent set. Then $\{v_1, \ldots, v_n\}$ is a basis of V.

- If $\dim V = n$ and $\{v_1, \ldots, v_k\}$ is a linearly independent set ($k < n$). Then one can find elements $v_{k+1}, \ldots, v_n$ such that $\{v_1, \ldots, v_n\}$ is a basis of V.

- If $\dim V = n$ and W is a subspace of V. Then $\dim W \leq \dim V$.

Proof. Exercise. □

From Monday:

Theorem. *Suppose V is a v.s., $\{v_1, \ldots, v_m\}$ is a basis of size m, and $w_1, \ldots, w_n \in V$ (all distinct) with $n > m$. Then $\{w_1, \ldots, w_n\}$ is linearly dependent.*

Corollary. *If V has one basis with m elements and another basis with n elements, then $m = n$.*

Question: can V have one basis with m elements and another basis with infinitely many elements?
Answer: no; else the infinite basis of V would have a subset of size $m+1$, which is automatically linearly independent.

Corollary. *If V is finitely spanned, then any two bases have the same (finite) number of elements.*

In this case, $\dim V$ is by definition the (finite) number of elements in any basis.

Fact: Even if V is not finitely spanned, any two bases for V have the same cardinality. (This is not easy to prove.)

In this course we will simply write $\dim V = \infty$ if V has an infinite basis (equivalently, V has no finite basis). So e.g., $\dim \mathbb{R}[x] = \infty = \dim \mathbb{R}^{[0,1]}$. But warning: in advanced linear algebra, we do not write $\dim \mathbb{R}[x] = \dim \mathbb{R}^{[0,1]}$, because actually

$$\dim \mathbb{R}[x] = \aleph_0 \qquad \text{while} \qquad \dim \mathbb{R}^{[0,1]} = 2^{2^{\aleph_0}}.$$

Corollary. *If V is finitely spanned and $\mathcal{B} = \{w_1, \ldots, w_n\}$ is linearly independent, then $\mathcal{B}$ can be extended to a basis for V. I.e., $\exists v_1, \ldots, v_r$ such that $\{w_1, \ldots, w_n, v_1, \ldots, v_r\}$ is a basis for V.*

The proof idea is simple: either $\mathcal{B}$ is already a basis, or else span$\mathcal{B} \subset V$. In the latter case, choose any $v_1 \in V \setminus \text{span}(\mathcal{B})$. Then $\mathcal{B} \cup \{v_1\}$ is linearly independent (by Jan. 15 thm).
Repeat. This can't go on forever because linearly independent sets must have size $\leq \dim V$. □

Fact: Even if V is not finitely spanned, every linearly independent subset of V can be extended to a basis (this is proved using some form of the Axiom of Choice).

Recall: a finite linearly independent set $\{v_1, \ldots, v_n\}$ in a v.s. V is a *maximal linearly independent set* if for every $w \in V \setminus \{v_1, \ldots, v_n\}$, $\{v_1, \ldots, v_n, w\}$ is linearly dependent.

Corollary. *In a finitely spanned vector space, every maximal linearly independent set is a basis (and vice versa).*

Definition. More generally, a subset $\mathcal{B} \subseteq V$ is a *maximal linearly independent set* if it is linearly independent and for all $x \in V \setminus \mathcal{B}$, $\mathcal{B} \cup \{w\}$ is linearly dependent.

Fact: Even if V is not finitely spanned, every maximal linearly independent set is a basis and vice versa. (Exercise)

We can also "shrink" spanning sets to bases.

Fact: In any vector space V, if $\mathcal{B} \subseteq V$ and span$\mathcal{B} = V$, then there exists a subset $\mathcal{B}' \subseteq \mathcal{B}$ such that $\mathcal{B}'$ is a basis for V.

We proved this fact for countably spanned vector spaces. One needs to use the Axiom of Choice to prove it in general.

Definition. A subset $\mathcal{B} \subseteq V$ is a *minimal spanning subset of V* if span$\mathcal{B} = V$ and for every $w \in \mathcal{B}$, span$(\mathcal{B} \setminus \{w\}) \neq V$.

Fact: In any vector space, every minimal spanning set is a basis and vice versa.

Definition 1. Suppose V is a vector space and W is a subspace. If $x, y \in V$, we write $x \equiv y \pmod W$ if $x - y \in W$.

Claim 2. *In the above setting,* $\equiv \pmod W$ *is an equivalence relation on* V.

Definition 3. In the above setting, given $x \in V$, we let $x + W \overset{df}{=} \{x + w : w \in W\}$, and call $x + W$ the *translation of* W *by* x (or the *coset of* W *containing* x).

(Note that x is fixed: $x + W$ is the set gotten by adding x to all possible vectors in W. For an example, think of W being a line through the origin; then $x + W$ is the line parallel to W going through x.)

Lemma 4. *If* V *is a vector space and* W *is a subspace, then for any* $x \in V$*, the equivalence class of* $\equiv \pmod W$ *containing* x *is exactly* $x + W$.

Corollary. *With* V *and* W *as above, for any* $x, y \in V$,

$$x + W = y + W \iff x \equiv y \pmod W \quad \textit{i.e.}, \ x - y \in W.$$

Definition 5. Given a vector space V and a subspace W, V/W denotes the set of all translations of W. Formally, $V/W = \{x + W : s \in V\}$.

Definition 6. Let V be a vector space over $\mathbb{F}$, and let W be a subspace. Operations of addition and scalar-multiplication-by-$\mathbb{F}$ are defined naturally on V/W by representatives:

$$\begin{aligned} (x + W) + (y + W) &:= (x + y) + W \\ c(x + W) &:= (cx) + W. \end{aligned}$$

Remark: the middle $+$ in the expression $(x + W) + (y + W)$ is not the same operation as the other $+$'s. It is an operation on (certain) sets. If $S = x + W$ and $T = y + W$, then the definition of $S + T$ is not $\{s + t : s \in S \text{ and } t \in T\}$. The definition is: choose x, y so that $S = x + W$ and $T = y + W$; add x and y to get $x + y = z$; then $S + T$ is defined to be the translation of W through z (i.e., $z + W$).

Claim 7. *In the above situation,*

1. *The two operations are well-defined, and*
2. *The set* V/W *with these operations is a vector space over* $\mathbb{F}$.

(1) means the following: if $x + W = x_1 + W$ and $y + W = y_1 + W$, then $(x + y) + W$ should equal $(x_1 + y_1) + W$, and $(cx) + W$ should equal $(cx_1) + W$ for all $c \in \mathbb{F}$. Here is a proof of the second part:

$$\begin{aligned} x + W = x_1 + W &\implies x - x_1 \in W \\ &\implies c(x - x_1) \in W \\ &\implies (cx) - (cx_1) \in W \\ &\implies (cx) + W = (cx_1) + W. \end{aligned}$$

(2) means that the set V/W with the operations given above satisfies axiom (VS 1) – (VS 8). This is an exercise. (What is the "zero vector"?)

Definition. V/W with the natural operations is called the *quotient space of* V *modulo* W.

The next definition describes the "good" functions between vector spaces.

Definition. Let V and W be vector spaces over the same field $\mathbb{F}$. A function $T : V \to W$ is called a **linear transformation**, or is said to be **linear**, if:

(1) $T(x + y) = T(x) + T(y)$ for all $x, y \in V$, and
(2) $T(ax) = aT(x)$ for all $x \in V$ and $a \in \mathbb{F}$.

Example 2.1.

(1) ($\mathbb{F} = \mathbb{R}$). Let $V = W = \mathbb{R}$. Fix $\lambda \in \mathbb{R}$. Define $T : \mathbb{R} \to \mathbb{R}$ by $T(x) = \lambda x$. **Claim:** T is a linear.
Remark: every linear transformation $\mathbb{R} \to \mathbb{R}$ has this form (for some λ).

(2) ($\mathbb{F} = \mathbb{R}$). Let $V = W = \mathbb{R}^2$. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x_1, x_2) = (-x_2, x_1)$. (This is just "rotation c.c.w. by $90°$ about $(0, 0)$.")

Claim. *T is a linear transformation.*

(3) The previous example is a special case of the following: let $A \in M_{m\times n}(\mathbb{R})$, say

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Given $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^m.$$

Then define $L_A : \mathbb{R}^n \to \mathbb{R}^m$ by $L_A(x) = Ax$. **Claim:** L_A is linear.

Can generalize this further: replace $\mathbb{R}$ with any field $\mathbb{F}$. Given $A \in M_{m\times n}(\mathbb{F})$, get a linear transformation $L_A : \mathbb{F}^n \to \mathbb{F}^m$. We'll see later that every linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ has this form.

(4) Define $T : C([-1, 1]) \to \mathbb{R}$ by $T(f) = \int_{-1}^{1} f(x)dx$. **Claim:** T is linear.
(5) Define $D : C^1(\mathbb{R}) \to C(\mathbb{R})$ by $D(f) = f'$. **Claim:** D is linear.

Here are some easily proved properties of all linear transformations $T : V \to W$.

(1) $T(0) = 0$.
(2) $T(x - y) = T(x) - T(y)$.
(3) $T(a_1x_1 + \cdots + a_nx_n) = a_1T(x_1) + \cdots + a_nT(x_n)$.

Example 2.1 (Continued)**.**

(5) $T : M_{m\times n}(\mathbb{F}) \to M_{n\times m}(\mathbb{F})$ given by $T(A) = A^t$ (transpose).
(6) Given any V and W, the function $T_0 : V \to W$ which maps every $x \in V$ to the 0 vector in W. (Called the **zero transformation**.)
(7) Given any V, the function $I_V : V \to V$ defined by $I_V(x) = x$ for all $x \in V$. (This is the **identity function** on V.)

Definition. Suppose $T : V \to W$ is linear.

(1) The **null space** of T is the set $N(T) = \{x \in V \,:\, T(x) = 0\}$.
(2) The **range** of T is the set $R(T) = \{T(x) \,:\, x \in V\}$.

Note that $N(T) \subseteq V$ and $R(T) \subseteq W$.

Example. Define $D_n : \mathsf{P}_n(\mathbb{R}) \to \mathsf{P}_n(\mathbb{R})$ by $D_n(f) = f'$. Obviously D_n is linear.

(1) $N(D_n) = \{f \in \mathsf{P}_n(\mathbb{R}) \,:\, f' = 0\} = \{\text{constant polynomials}\} = \operatorname{span}(1)$.
(2) $R(D_n) = \{f' \,:\, f \in \mathsf{P}_n(\mathbb{R})\} = \mathsf{P}_{n-1}(\mathbb{R})$.

Theorem. *Let $T : V \to W$ be linear. Then $N(T)$ is a subspace of V, and $R(T)$ is a subspace of W.*

Because linear transformations preserve linear combinations, we can prove the following.

Theorem (Useful Trick Theorem)**.** *Suppose $T : V \to W$ is linear and $V = \operatorname{span}(v_1, \dots, v_n)$. Then*

$$R(T) = \operatorname{span}(T(v_1), \dots, T(v_n)).$$

Example. Let $A \in M_{m\times n}(\mathbb{F})$ and consider $L_A : \mathbb{F}^n \to \mathbb{F}^m$. $\mathbb{F}^n$ is spanned by $\{e_1, e_2, \dots, e_n\}$ where

$$e_i \;=\; (0, \dots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \dots, 0).$$

Thus $R(L_A) = \operatorname{span}(L_A(e_1), \dots, L_A(e_n))$ by the Useful Trick Theorem. Note that

$$L_A(e_i) \;=\; \begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \;=\; \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} \;=\; \text{the } i\text{th column of } A.$$

Hence $R(L_A)$ is the subspace of $\mathbb{F}^m$ spanned by the columns of A.

The Useful Trick Theorem can be helpful in deciding whether a linear transformation $T : V \to W$ is **surjective**, i.e., satisfies $R(T) = W$.

The next theorem gives a useful test for deciding whether T is **injective**.

Theorem. *A linear transformation $T : V \to W$ is injective iff $N(T) = \{0\}$.*

Announcements:

(1) Monday's Tutorial: required material

Definition. A linear transformation $T : V \to W$ is an *isomorphism* if it is injective and surjective (i.e., bijective). When this happens, we also write $T : V \cong W$.

We write $V \cong W$ and say V *is isomorphic to* W, if there exists an isomorphism $T : V \cong W$.

Example: $\mathsf{P}_n(\mathbb{R}) \cong \mathbb{R}^{n+1}$. One isomorphism is

$$T(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = (a_0, a_1, \ldots, a_n).$$

Definition. Suppose $T : V \to W$ is linear. Then $\operatorname{rank}(T) := \dim(R(T))$ and $\operatorname{nullity}(T) = \dim(N(T))$.

Theorem (Rank-Nullity Theorem)**.** *Suppose $T : V \to W$ is linear and $\dim(V) < \infty$. Then $\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V)$.*

Proof sketch. Suppose $\dim(V) = n$ and $\dim(N(T)) = k \leq n$. Let $m = n - k$. Must show $\dim(R(T)) = m$.

Pick a basis $S = \{v_1 \ldots, v_k\}$ for $N(T)$. By A2Q2, S can be extended to a basis $B \supseteq S$ for V. Thus $|B| = n = k + m$, so we can write $B = \underbrace{\{v_1, \ldots, v_k\}}_{S} \cup \{x_1, \ldots, x_m\}$.

Let $C = \{T(x_1), \ldots, T(x_m)\}$. It suffices to show that $|C| = m$ and C is a basis for $R(T)$. Then it will follow that $\dim(R(T)) = m$. $\square$

Consider the proof in the special case when $T : V \cong W$. Then $N(T) = \{0\}$, so $S = \varnothing$, so $m = n$ and $\{x_1, \ldots, x_n\}$ is a basis for V. The proof shows that $\{T(x_1), \ldots, T(x_n)\}$ is a basis for $R(T)$, which is W. This proves:

Corollary. *If $T : V \cong W$ and $\dim(V) < \infty$, then $\dim(V) = \dim(W)$ and T sends any basis of T to a basis of W.*

Here is another cute consequence of the RNT.

Corollary. *Suppose $T : V \to W$ is linear and $\dim(V) = \dim(W) = n < \infty$. Then T is injective iff T is surjective.*

Proof. Observe that

(1) T is injective $\iff$ $N(T) = \{0\}$ $\iff$ $\operatorname{nullity}(T) = 0$, and
(2) T is surjective $\iff$ $R(T) = W$ $\iff$ $\operatorname{rank}(T) = n$.

Since $\operatorname{nullity}(T) + \operatorname{rank}(T) = n$, the claim holds. $\square$

Proposition. *Suppose $\{v_1, \ldots, v_n\}$ is a basis for vector space V over $\mathbb{F}$. Then for every $x \in V$, x can be written* uniquely *as*
$$x = a_1 v_1 + \cdots + a_n v_n, \quad a_1, \ldots, a_n \in \mathbb{F}.$$

Proof sketch. (In class) □

Example: $W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$ with basis $\{v_1, v_2\}$ where $v_1 = (-1, 1, 0)$ and $v_2 = (0, -1, 1)$. (Visualizing points in $v \in W$ via the pair of numbers (a, b) such that $v = av_1 + bv_2$.)

Definition. Let V be a finite-dimensional vector space. An **ordered basis** for V is a basis $(v_1, \ldots, v_n)$, ordered as an n-tuple.

Following the text, I'll use α, β, γ etc for ordered bases.

Definition. Given a vector space V over $\mathbb{F}$ with $\dim(V) = n$, an ordered basis $\beta = (v_1, \ldots, v_n)$ for V, and a vector $x \in V$, the **coordinate vector of x relative to β** is the unique n-tuple $(a_1, \ldots, a_n) \in \mathbb{F}^n$ satisfying
$$x = a_1 v_1 + \cdots + a_n v_n.$$
We denote $(a_1, \ldots, a_n)$ by $[x]_\beta$.

Example 2.15. In the previous example, let $\beta = (v_1, v_2)$ where $v_1 = (-1, 1, 0)$ and $v_2 = (0, -1, 1)$. If $x = (-3, 1, 2)$ then $[x]_\beta = (3, 2)$.

If β is an ordered basis for V (where V is a finite-dimensional vector space over $\mathbb{F}$ with $n = \dim(V)$), then we can view $[\ \ ]_\beta$ as a function $V \to \mathbb{F}^n$.

Theorem 2.16. *Let V be a finite-dimensional vector space over $\mathbb{F}$, $\dim(V) = n$, and let β be an ordered basis for V. The map $[\ \ ]_\beta : V \to \mathbb{F}^n$ is a bijective linear transformation (i.e., an isomorphism).*

Proof. (In class) □

Proposition. *Suppose V, W are vector spaces over $\mathbb{F}$, B is a basis for V, and $T : V \to W$ is a linear transformation. Then T is completely determined by its values $T(v)$, $v \in B$.*

I.e., if $T' : V \to W$ is another linear transformation and $T(v) = T'(v)$ for all $v \in B$, then $T = T'$.

Proof #1. Let T' be another linear transformation with $T|_B = T'|_B$. Let $x \in V$. x can be written

$$x = a_1 v_1 + \cdots + a_n v_n$$

for some $v_1, \ldots v_n \in B$ and $a_1, \ldots, a_n \in \mathbb{F}$. Then

$$T'(x) = T'(a_1 v_1 + \cdots + a_n v_n) = a_1 T'(v_1) + \cdots + a_n T'(v_n) = \cdots = T(x).$$

Since $x \in V$ was arbitrary, $T = T'$. □

Proof #2. Claim: the set of all linear transformations from V to W is a subspace of W^V. (Exercise). This set is called $\mathrm{Hom}(V, W)$. Now define $D = T - T'$, i.e., $D(x) = T(x) - T'(x)$. $T, T' \in \mathrm{Hom}(V, W)$ so $D \in \mathrm{Hom}(V, W)$, meaning D is linear. Let's prove D is constantly 0 by showing $N(D) = V$. By hypothesis, $B \subseteq N(D)$. Since $N(D)$ is a subspace of V, we get $\mathrm{span}(B) \subseteq N(D)$, i.e, $V \subseteq N(D)$. □

Proposition. *Suppose V, W, B are as above. Every function $\tau : B \to W$ extends (uniquely) to a linear transformation $T : V \to W$, i.e., with $T|_B = \tau$.*

We call this "freely extending" τ.

Proof. First we say how to define T. Given $x \in V$, x can be written

$$x = a_1 v_1 + \cdots + a_n v_n$$

with $v_1, \ldots, v_n \in B$ and $a_1, \ldots, a_n \in \mathbb{F}$. Define

$$T(x) := a_1 \tau(v_1) + \cdots + a_n \tau(v_n).$$

Now show that T extends τ and is linear (done in class). □

Example. Let $V = \mathbb{R}^3$ and $W = \mathbb{R}^2$. Choose the basis $\{v_1, v_2, v_3\}$ where

$$v_1 = (1, 0, 1), \qquad v_2 = (1, 0, -1), \qquad v_3 = (1, 1, 1).$$

Ask for three random vectors $A, B, C \in \mathbb{R}^2$. Define $\tau : \{v_1, v_2, v_3\} \to \mathbb{R}^2$ by $\tau(v_1) = A$, $\tau(v_2) = B$, and $\tau(v_3) = C$. Now find the unique linear transformation $\mathbb{R}^3 \to \mathbb{R}^2$, of the form L_A (where $A \in M_{2\times 3}(\mathbb{R})$), extending τ.

Example. Let V be a vector space over $\mathbb{F}$ and $\dim(V) = n$. Let $\beta = (v_1, \ldots, v_n)$ be an ordered basis for V. Define $\tau : \{v_1, \ldots, v_n\} \to \mathbb{F}^n$ by $\tau(v_i) = e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ for $i = 1, \ldots, n$. τ freely extends to a (unique) linear transformation $T : V \to \mathbb{F}^n$, namely, to $T = [\ \ ]_\beta$.

Example. Let $V, \mathbb{F}, \beta$ be as before. Pick $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{F}^n$. Define $\tau_{\mathbf{a}} : \{v_1, \ldots, v_n\} \to \mathbb{F}$ by $\tau_{\mathbf{a}}(v_i) = a_i$ for $i = 1, \ldots, n$. $\tau_{\mathbf{a}}$ extends to a (unique) linear transformation $f_{\mathbf{a}} : V \to \mathbb{F}$ (i.e., to a *linear functional*), which satisfies

$$f_{\mathbf{a}}(c_1 v_1 + \cdots + c_n v_n) = c_1 a_1 + \cdots + c_n a_n.$$

In particular, if $\mathbf{a} = e_1 = (1, 0, 0, \ldots, 0)$ then

$$f_{e_1}(c_1 v_1 + \cdots + c_n v_n) = c_1.$$

In other words, f_{e_1} is the linear functional f_1 in the dual basis $\beta^* = (f_1, \ldots, f_n)$ for V^* defined in Monday's tutorial. Similarly, $f_{e_i} = f_i$ for $i = 2, \ldots, n$.

Announcements:

(1) No tutorial on Monday Feb 10.

Suppose $T : V \to W$ is linear where V and W are both finite-dimensional vector spaces over $\mathbb{F}$. Suppose $\beta = (v_1, \dots, v_n)$ is an ordered basis for V and $\gamma = (w_1, \dots, w_m)$ is an ordered basis for W.

- T is completely determined by $T(v_1), \dots, T(v_n)$.
- Each $T(v_j)$ is determined by its coordinate vector $[T(v_j)]_\gamma \in \mathbb{F}^m$.

Definition. In this context, the **matrix representation of T for β and γ** is the matrix $A \in M_{m\times n}(\mathbb{F})$ whose *columns* are $[T(v_1)]_\gamma, \dots, [T(v_n)]_\gamma$. Thus

$$A \;=\; \begin{pmatrix} \big| & \cdots & \big| & \cdots & \big| \\ [T(v_1)]_\gamma & \cdots & [T(v_j)]_\gamma & \cdots & [T(v_n)]_\gamma \\ \big| & \cdots & \big| & \cdots & \big| \end{pmatrix}.$$

We denote this matrix A by $[T]_\beta^\gamma$.

Example. Let $A \in M_{m\times n}(\mathbb{F})$ and $T = L_A : \mathbb{F}^n \to \mathbb{F}^m$, where $L_A(x) = Ax$. Let σ_n be the *standard ordered basis for* $\mathbb{F}^n$, i.e., $\sigma_n = (e_1, \dots, e_n)$ where $e_j = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{F}^n$ with 1 in the j-th position. Similarly let σ_m be the standard ordered basis for $\mathbb{F}^m$.

Claim: $[L_A]_{\sigma_n}^{\sigma_m} = A$ (not surprisingly).

Proof. First recall that, for each $j = 1, \dots, n$, Ae_j equals the j-th column of A. Second, note that for any $x = (b_1, \dots, b_m) \in \mathbb{F}^m$ we have $x = b_1e_1 + \cdots + b_me_m$ and so $[x]_{\sigma_m} = (b_1, \dots, b_m) = x$.

Now the j-th column of $[L_A]_{\sigma_n}^{\sigma_m}$ is by definition

$$\begin{aligned} [L_A(e_j)]_{\sigma_m} &= [Ae_j]_{\sigma_m} \\ &= Ae_j && \text{by the 2nd remark} \\ &= \text{the } j\text{-th column of } A && \text{by the 1st remark.} \end{aligned}$$

Thus $[L_A]_{\sigma_n}^{\sigma_m}$ and A have the same columns, so are the same matrix. $\square$

Theorem 2.21. *Suppose V, W are finite-dimensional vector spaces over $\mathbb{F}$, $\beta = (v_1, \dots, v_n)$ is an ordered basis for V, $\gamma = (w_1, \dots, w_m)$ is an ordered basis for W, and $T : V \to W$ is linear. Then for all $x \in V$,*

$$[T(x)]_\gamma \;=\; [T]_\beta^\gamma \cdot [x]_\beta.$$

Proof. Write

$$[T]_\beta^\gamma \;=\; \begin{pmatrix} \big| & \cdots & \big| & \cdots & \big| \\ [T(v_1)]_\gamma & \cdots & [T(v_j)]_\gamma & \cdots & [T(v_n)]_\gamma \\ \big| & \cdots & \big| & \cdots & \big| \end{pmatrix} \;=\; \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

meaning

$$T(v_j) \;=\; a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m \qquad \text{for } j = 1, \dots, n.$$

Also write

$$[x]_\beta \;=\; (c_1, \dots, c_n),$$

meaning

$$x = c_1 v_1 + \cdots + c_n v_n.$$

On the one hand,

$$[T]_\beta^\gamma \cdot [x]_\beta = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_n \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2n}c_n \\ \vdots \\ a_{m1}c_1 + a_{m2}c_2 + \cdots + a_{mn}c_n \end{pmatrix}.$$

On the other hand,

$$\begin{aligned} T(x) &= T(c_1 v_1 + \cdots + c_n v_n) \\ &= c_1 T(v_1) + \cdots + c_n T(v_n) \qquad (T \text{ is linear}) \\ &= c_1(a_{11}w_1 + \cdots + a_{m1}w_m) + \cdots + c_n(a_{1n}w_1 + \cdots + a_{mn}w_m) \\ &= (c_1 a_{11} + \cdots + c_n a_{1n})w_1 + \cdots + (c_1 a_{m1} + \cdots + c_n a_{mn})w_m \\ &= (a_{11}c_1 + \cdots + a_{1n}c_n)w_1 + \cdots + (a_{m1}c_1 + \cdots + a_{mn}c_n)w_m. \end{aligned}$$

Hence we can see that $[T]_\beta^\gamma \cdot [x]_\beta$ is the coordinate vector of $T(x)$ relative to γ, proving the theorem. $\square$

Notation. Suppose $A \in M_{m\times n}(\mathbb{F})$.

- For $j = 1, \ldots, n$, $\mathrm{Col}_j(A)$ denotes the j-th column of A. Thus $\mathrm{Col}_j(A) \in \mathbb{F}^m$.
- For $i = 1, \ldots, m$, $\mathrm{Row}_i(A)$ denotes the i-th row of A. Thus $\mathrm{Row}_i(A) \in \mathbb{F}^n$.

I may write

$$A = \left(\boxed{\begin{matrix} | \\ \mathrm{Col}_1 \\ | \end{matrix}}, \ldots, \boxed{\begin{matrix} | \\ \mathrm{Col}_n \\ | \end{matrix}} \right) = \left(\begin{matrix} \boxed{-\ \mathrm{Row}_1\ -} \\ \vdots \\ \boxed{-\ \mathrm{Row}_m\ -} \end{matrix} \right).$$

Recall that A determines a linear transformation $L_A : \mathbb{F}^n \to \mathbb{F}^m$ defined by $L_A(x) = Ax$. Also recall:

(1) $Ae_j = \mathrm{Col}_j(A)$ for each $j = 1, \ldots, n$.

(2) Whenever $T : V \to W$ is a linear transformation, $\alpha = (v_1, \ldots, v_n)$ is an ordered basis for V, and $\beta = (w_1, \ldots, w_m)$ is an ordered basis for W, then $[T]_\alpha^\beta$ is the $m \times n$ matrix whose j-th column is given by $\mathrm{Col}_j([T]_\alpha^\beta) = [T(v_j)]_\beta$ for $j = 1, \ldots, n$.

With this information we can calculate $[L_A]_{\sigma_n}^{\sigma_m}$; it is the $m \times n$ matrix whose j-th column is

$$\mathrm{Col}_j([L_A]_{\sigma_n}^{\sigma_m}) = [L_A(e_j)]_{\sigma_m} = [Ae_j]_{\sigma_m} = [\mathrm{Col}_j(A)]_{\sigma_m} = \mathrm{Col}_j(A),$$

where the last equality is because $\mathrm{Col}_j(A) \in \mathbb{F}^m$ and σ_m is the standard ordered basis for $\mathbb{F}^m$. This proves that $[L_A]_{\sigma_n}^{\sigma_m} = A$; that is, the matrix representation of L_A with respect to the standard ordered bases is A.

Definition. Let $\mathbb{F}$ be a field. Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$. The **matrix product** AB is the $m \times p$ matrix $C \in M_{m\times p}(\mathbb{F})$ whose row-i, column-j entry is the linear combination of the entries in $\mathrm{Col}_j(B)$ using as scalars the entries of $\mathrm{Row}_i(A)$. That is,

$$\begin{pmatrix} a_{11} & \cdots & a_{1t} & \cdots & a_{1n} \\ & & \vdots & & \\ \boxed{a_{i1}} & \boxed{\cdots} & \boxed{a_{it}} & \boxed{\cdots} & \boxed{a_{in}} \\ & & \vdots & & \\ a_{m1} & \cdots & a_{mt} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & & \boxed{b_{1j}} & & b_{1p} \\ \vdots & & \boxed{\vdots} & & \vdots \\ b_{t1} & \cdots & \boxed{b_{tj}} & \cdots & b_{tp} \\ \vdots & & \boxed{\vdots} & & \vdots \\ b_{n1} & & \boxed{b_{nj}} & & b_{np} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1p} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & \boxed{c_{ij}} & \cdots & c_{ip} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mj} & \cdots & c_{mp} \end{pmatrix}$$

where each entry c_{ij} of the product is given by $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

Remarks.

(1) If $p = 1$, so B and AB are column vectors, then the definition above is just our usual definition for multiplying a matrix by a column vector.

(2) In general (i.e., when B has several columns), B and AB have the same number of columns, and the j-th column of AB is obtained by multiplying A by the j-th column of B. I.e., $\mathrm{Col}_j(AB) = A \cdot \mathrm{Col}_j(B)$ for $j = 1, \ldots, p$.

Now suppose we have $T : V \to W$ and $U : W \to Z$, both linear, where V, W, Z are all finite-dimensional, say $\dim(V) = p$, $\dim(W) = n$, and $\dim(Z) = m$. In this situation, define $UT \overset{\text{def}}{=} U \circ T : V \to Z$; it is also linear (exercise). Also assume that α, β, γ are ordered bases for V, W, Z respectively.

Theorem 2.22. *In this situation,* $[UT]_\alpha^\gamma = [U]_\beta^\gamma \cdot [T]_\alpha^\beta$.

Proof. Given in class. (Show $\mathrm{Col}_j(LHS) = \mathrm{Col}_j(RHS)$ for $j = 1, \ldots, p$.) $\square$

Here is an application of Monday's Theorem. Let $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$. Thus $L_B : \mathbb{F}^p \to \mathbb{F}^n$ and $L_A : \mathbb{F}^n \to \mathbb{F}^m$. So we can compose L_B with L_A to get $L_A L_B : \mathbb{F}^p \to \mathbb{F}^m$.

Corollary. *In this situation, $L_A L_B = L_{AB}$.*

Proof. It suffices to show that $[L_A L_B]_{\sigma_p}^{\sigma_m} = [L_{AB}]_{\sigma_p}^{\sigma_m}$, since linear transformations are determined by their matrix representations. Apply Monday's Theorem to the LHS and use $[L_D]_{\sigma_n}^{\sigma_m} = D$ repeatedly. □

Corollary. *Matrix multiplication (when defined) is associative. I.e., if $A \in M_{m\times n}(\mathbb{F})$, $B \in M_{n\times p}(\mathbb{F})$, and $C \in M_{p\times r}(\mathbb{F})$, then $(AB)C = A(BC)$.*

Proof. It suffices to show $L_{(AB)C} = L_{A(BC)}$ (as functions). Use the previous Corollary and the fact that composition of functions is associative. □

Warning: Matrix multiplication is not in general commutative.

Definition. A square matrix $A \in M_{n\times n}(\mathbb{F})$ is **invertible** if there exists $B \in M_{n\times n}(\mathbb{F})$ satisfying $AB = BA = I_n$.

Note that if such B exists, then B is unique. (Proof: if B_1 and B_2 satisfy $AB_1 = B_1 A = I_n$ and $AB_2 = B_2 A = I_n$, then $B_1 = B_1 I_n = B_1(AB_2) = (B_1 A)B_2 = I_n B_2 = B_2$.) This justifies the following:

Notation. If A is invertible, then the unique matrix B satisfying $AB = BA = I_n$ is denoted A^{-1} and is called the **inverse** of A.

Theorem 2.24. *Suppose V, W are fin. dim. vector spaces over $\mathbb{F}$, α, β are ordered bases for V, W respectively, and $T : V \to W$ is linear. T is an isomorphism iff $[T]_\alpha^\beta$ is invertible, in which case $([T]_\alpha^\beta)^{-1} = [T^{-1}]_\beta^\alpha$.*

Proof. ($\Rightarrow$) Let $A = [T]_\alpha^\beta$. Assume that T is an isomorphism. Then $\dim(V) = \dim(W) = n$, say (by a corollary of the Rank-Nullity Theorem; see Jan. 31 lecture), so A is a square ($n \times n$) matrix. Let $T^{-1} : W \to V$ be the inverse linear transformation to T. Then $B := [T^{-1}]_\beta^\alpha$ is also an $n \times n$ matrix, and

$$\begin{aligned} AB &= [T]_\alpha^\beta \cdot [T^{-1}]_\beta^\alpha \\ &= [TT^{-1}]_\beta^\beta && \text{Monday's Theorem} \\ &= [I_W]_\beta^\beta \\ &= I_n && \text{(exercise).} \end{aligned}$$

A similar proof shows $BA = [I_V]_\alpha^\alpha = I_n$. So by definition, A is invertible $A^{-1} = B$.

($\Leftarrow$) exercise. □

Easy Lemma. *If $A, B \in M_{n\times n}(\mathbb{F})$ are invertible, then AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. Let $C = B^{-1}A^{-1}$. It suffices to show that $(AB)C = C(AB) = I_n$, for then it will follow that AB is invertible and its inverse is C. So let's check:

$$(AB)C = (AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

The proof of $C(AB) = I_n$ is similar. □

Recall the following (Corollary of Rank-Nullity Theorem, Jan. 31):

Corollary. *Suppose $T : V \to W$ is linear and $\dim(V) = \dim(W) = n < \infty$. Then T is injective $\iff$ T is surjective $\iff$ T is an $\cong$.*

Fact. Suppose $f : X \to Y$ and $g : Y \to Z$, so $gf := g \circ f : X \to Z$. Assume gf is a bijection. Then f is injective and g is surjective.

Proof. Exercise. □

Theorem. *Suppose $A, B \in M_{n\times n}(\mathbb{F})$. If AB is invertible, then A and B are both invertible.*

Proof. AB invertible $\implies$ $L_{AB} : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism.
$\implies$ $L_A L_B : \mathbb{F}^n \to \mathbb{F}^n$ is an isomorphism (hence bijection), because $L_{AB} = L_A L_B$ (Feb. 12).
$\implies$ L_B is injective and L_A is surjective (today's Fact).
$\implies$ L_A and L_B are isomorphisms (Jan. 31 corollary).
$\implies$ A and B are invertible. □

Corollary. *If A, B are $n \times n$ matrices and $AB = I_n$, then $BA = I_n$ (so A is invertible and $B = A^{-1}$).*

Proof. Since $AB = I_n$ and I_n is invertible, the Theorem gives that A is invertible, so A^{-1} exists. Multiplying $AB = I_n$ on the left by A^{-1} gives $B = A^{-1}$. So of course $BA = I_n$. □

Now back to discussing coordinatization.

Suppose V is a fin. dim. vector space over $\mathbb{F}$. Suppose β and γ are two ordered bases for V and $x \in V$. If we know $[x]_\beta$, how can we find $[x]_\gamma$ and vice versa?

Theorem. *In this situation, let $Q = [I_V]_\beta^\gamma$. Then:*

(1) *Q is invertible.*
(2) *For any $x \in V$, $Q[x]_\beta = [x]_\gamma$ and $Q^{-1}[x]_\gamma = [x]_\beta$.*

Proof. (1) I_V is an isomorphism, so Q is invertible by a Theorem from Wednesday.

(2) $Q[x]_\beta = [I_V]_\beta^\gamma \cdot [x]_\beta = [I_V(x)]_\gamma$ by the Theorem from Feb. 7. But this equals $[x]_\gamma$. Now multiply this equation on the left by Q^{-1} to get $[x]_\beta = Q^{-1}[x]_\gamma$. □

Definition. The matrix $Q = [I_V]_\beta^\gamma$ is called the **change of coordinates matrix from β to γ**.

Definition. If V is fin. dim., β is an ordered basis for V, and $T : V \to V$ is linear, then $[T]_\beta$ denotes $[T]_\beta^\beta$.

Theorem. *Suppose V is finite-dimensional, $T : V \to V$ is linear, and β, γ are two ordered bases for V. Let $Q = [I_V]_\beta^\gamma$ be the change of coordinates matrix from β to γ. Then*

$$[T]_\beta \;=\; Q^{-1}[T]_\gamma Q.$$

Proof. It suffices to prove $Q[T]_\beta = [T]_\gamma Q$. To prove this, apply the Theorem from Feb 10 to show each side equals $[T]_\beta^\gamma$. □

Recall the following facts from the Feb 10 lecture:

(1) $\text{Col}_j(AB) = A \cdot \text{Col}_j(B)$ for all $j = 1, \dots, n$ (if B has n columns)
(2) $Ae_j = \text{Col}_j(A)$ for $j = 1, \dots, n$ (if A has n columns, $e_j \in \mathbb{F}^n$).
(3) $Ax = \sum_{j=1}^n x_j \text{Col}_j(A)$ (if A has n columns and $x \in \mathbb{F}^n$).

Here are the analogous facts for rows:

(4) $\text{Row}_i(AB) = \text{Row}_i(A) \cdot B$ for all $i = 1, \dots, m$ (if A has m rows).
(5) $(e_i)^t A = \text{Row}_i(A)$ for $i = 1, \dots, m$ (if A has m rows).
(6) $x^t A = \sum_{i=1}^m x_i \text{Row}_i(A)$ (if A has m rows and $x \in \mathbb{F}^m$).

Definition. Let $A \in M_{m \times n}(\mathbb{F})$. An **elementary row operation** is any one of the following actions, resulting in a new matrix A':

(1) Switching two rows of A. $\quad R_i \leftrightarrows R_j$
(2) Multiplying one row of A by a nonzero scalar. $\quad R_i \leftarrow aR_i$ $(a \neq 0)$
(3) Adding a scalar multiple of one row of A to another row of A. $\quad R_i \leftarrow R_i + aR_j$

An **elementary column operation** is any action of the above kinds, but with rows replaced by columns. An elementary row or column operation is an **elementary operation**. An elementary operation is **type 1**, **type 2** or **type 3** according to whether it is obtained by rule (1), (2) or (3).

Newton's 3rd Law of Operations. *To every elementary operation there is an equal and opposite elementary operation.*

For example, the operation $R_i \leftarrow R_i + aR_j$ is undone by $R_i \leftarrow R_i + (-a)R_j$.

Definition. An **elementary matrix** is an $n \times n$ matrix which can be obtained by applying one elementary operation to I_n. It is of type 1, 2 or 3 according to the type of the operation used.

Notation. If $\mathcal{O}$ is an elementary operation on $m \times n$ matrices, $A \in M_{m \times n}(\mathbb{F})$, and A' is the result of applying $\mathcal{O}$ to A, then we write $A \xrightarrow{\mathcal{O}} A'$.

Theorem 3.1. *Fix m, n and suppose that $\mathcal{O}$ is an elementary column operation on $m \times n$ matrices. Let E be the elementary matrix obtained by applying $\mathcal{O}$ to I_n.*

Then for all $A \in M_{m \times n}(\mathbb{F})$, if $A \xrightarrow{\mathcal{O}} A'$ then $A' = AE$.

Proof sketch. Let $A_j := \text{Col}_j(A)$ for $j = 1, \dots, n$, so we can write $A = [A_1 \ A_2 \ \cdots \ A_n]$. The columns of I_n are $e_1, \dots, e_n$, so we can write $I_n = [e_1 \ e_2 \ \cdots \ e_n]$. Now consider cases according the type of the column operation is $\mathcal{O}$, and use Facts (1) and (2) judiciously. For example:

Case 3: $\mathcal{O}$ is $C_i \leftarrow C_i + aC_j$.

Then $E = [e_1, \dots, e_i + ae_j, \dots, e_n]$ and the columns of AE are $Ae_1, \dots, A(e_i + ae_j), \dots, Ae_n$. Note that $A(e_i + ae_j) = Ae_i + aAe_j$ (by linearity of L_A). Thus the columns of AE are $A_1, \dots, A_i + aA_j, \dots, A_n$. In other words, AE is the result of adding a times column j of A to to column i of A, so $A \xrightarrow{\mathcal{O}} AE$. $\square$

Theorem 3.2. *Fix m, n and suppose that $\mathcal{O}$ is an elementary row operation on $m \times n$ matrices. Let E be the elementary matrix obtained by applying $\mathcal{O}$ to I_m.*

Then for all $A \in M_{m \times n}(\mathbb{F})$, the result of applying $\mathcal{O}$ to A is EA.

From Theorems 3.1 and 3.2 we can deduce:

Theorem 3.3. *Elementary matrices are invertible. Moreover, if E is an elementary matrix corresponding to the elementary operation $\mathcal{O}$, then E^{-1} is the elementary matrix corresponding to the "opposite" elementary operation to $\mathcal{O}$.*

Definition. Suppose A, B are matrices of the same size. We write $A \rightsquigarrow B$ if there exists a sequence of elementary row and/or column operations that transforms A to B.

Theorem 3.4. *For every matrix $A \in M_{m\times n}(\mathbb{F})$ there exists a matrix D of the form*

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

where $r \geq 0$ and O_1, O_2, O_3 are all-zero matrices, such that $A \rightsquigarrow D$.

Proof sketch. If A is all 0s, we're done. Otherwise, A has a nonzero entry, and using type-1 operations we can move it to the 1,1 position. By a type-2 operation, we can change it to 1. Then using type-3 operations, we can "clear" the remaining entries in the first row and column. Thus we have converted A to a matrix A' of the form

$$A' = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

Repeat: if B is all 0s we're done. Else we can move a nonzero entry of B to the 2,2 position of A'; make it equal 1; and then clear the rest of the 2nd row and column. Etc. $\square$

Theorem 3.5. *If $A \in M_{m\times n}(\mathbb{F})$ and $A \rightsquigarrow B$, then there exist invertible $P \in M_{m\times m}(\mathbb{F})$ and $Q \in M_{n\times n}(\mathbb{F})$ such that $B = PAQ$.*

Proof sketch. Suppose $A \rightsquigarrow B$ via a sequence of elementary operations. Let $\mathcal{O}_1, \ldots, \mathcal{O}_k$ be the elementary row operations used, in this order, and $\mathcal{O}'_1, \ldots, \mathcal{O}'_\ell$ the elementary column operations used. Let $E_1, \ldots, E_k$ be the $m \times m$ elementary matrices corresponding to the row operations, and let $E'_1, \ldots, E'_\ell$ be the $n \times n$ elementary matrices corresponding to the column operations. Then

$$B = \underbrace{E_k \cdots E_2 E_1}_{=:P} A \underbrace{E'_1 E'_2 \cdots E'_\ell}_{=:Q}.$$

Each elementary matrix is invertible (Feb 24), and products of invertible matrices are invertible (Feb 12), so P and Q are invertible. $\square$

Suppose some evil math professor asks you to find the matrices P, Q promised by Theorem 3.5. What do you do?

Option 1: To find P, first find $E_1, \ldots, E_k$ and then multiply them (in the correct order). For Q, find $E'_1, \ldots, E'_\ell$ and multiply them.

Option 2: To find P, just apply the elementary row operations $\mathcal{O}_1, \ldots, \mathcal{O}_k$ to I_m. The resulting matrix will be P. To find Q, just apply the elementary column operations $\mathcal{O}'_1, \ldots, \mathcal{O}'_\ell$ to I_n. The resulting matrix will be Q.

Why does Option 2 work? The answer is easy: applying $\mathcal{O}_1, \ldots, \mathcal{O}_k$ to I_m is the same as multiplying I_n on the left by $E_1, \ldots, E_k$, so the result will be $E_k \cdots E_2 E_1 I_m = E_k \cdots E_2 E_1 = P$. Similarly, applying $\mathcal{O}'_1, \ldots, \mathcal{O}'_\ell$ to I_n is the same as multiplying I_n on the right by $E'_1, \ldots, E'_\ell$, so the result will be $I_n E'_1 E'_2 \cdots E'_\ell = E'_1 E'_2 \cdots E'_\ell = Q$.

Corollary. *If* $A \rightsquigarrow B$, *then*
(1) $B \rightsquigarrow A$.
(2) $A^t \rightsquigarrow B^t$.

Definition. Let $A \in M_{m\times n}(\mathbb{F})$.
(1) We call $\mathrm{span}(\{\mathrm{Row}_1(A), \dots, \mathrm{Row}_m(A)\})$ the **row space** of A.
(2) Similarly, we call $\mathrm{span}(\{\mathrm{Col}_1(A), \dots, \mathrm{Col}_n(A)\})$ the **column space** of A.

Recall that $\mathrm{span}(\{\mathrm{Col}_1(A), \dots, \mathrm{Col}_n(A)\}) = R(L_A)$. So the column space of A equals the range of L_A.

Definition.
(3) The **null space** of A, denoted $N(A)$, is the null space of L_A. I.e.,

$$N(A) := N(L_A) = \{x \in \mathbb{F}^n : Ax = 0\}.$$

Thus:
- The column space of A is a subspace of $\mathbb{F}^m$.
- The row space and null space of A are subspaces of $\mathbb{F}^n$.

Also note that the row space of A is identical to the column space of A^t and vice versa.

Definition.
(1) The **rank** of A is defined by $\mathrm{rank}(A) := \dim(R(L_A))$, i.e., the dimension of the column space of A.
(2) The **nullity** of A is defined by $\mathrm{nullity}(A) := \mathrm{nullity}(L_A)$.

Note that $\mathrm{rank}(A) + \mathrm{nullity}(A) = \dim(\mathbb{F}^n) = n$ by the Rank-Nullity Theorem applied to L_A.

Theorem 1. *If* $A \in M_{m\times n}(\mathbb{F})$ *and* $Q \in M_{n\times n}$ *with* Q *invertible, then* $R(L_A) = R(L_{AQ})$.

Proof sketch. Consider $L_{AQ} = L_A \circ L_Q$. Note that L_Q is an isomorphism (because Q is invertible). □

Corollary 1. *If* $A \rightsquigarrow B$ *entirely by column operations, then* A *and* B *have the same column space.*

Corollary 2. *If* $A \rightsquigarrow B$ *entirely by row operations, then* A *and* B *have the same row space.*

Proof sketch. If $A \rightsquigarrow B$ by row operations, then $A^t \rightsquigarrow B^t$ by column operations. □

Lemma. *Suppose* V *is a finite-dimensional space,* $T : V \cong V'$, *and* W *is a subspace of* V. *Let* $W' = \{T(w) : w \in W\}$. *Then* $\dim(W) = \dim(W')$.

Proof sketch. Let $\{x_1, \dots, x_k\}$ be a basis for W. Prove that $\{T(x_1), \dots, T(x_k)\}$ is a basis for W'. □

Theorem 2. *Suppose* $A \in M_{m\times n}(\mathbb{F})$ *and* $P \in M_{m\times m}(\mathbb{F})$ *with* P *invertible. Then* $\mathrm{rank}(A) = \mathrm{rank}(PA)$.

Proof sketch. Consider $L_{PA} = L_P \circ L_A$. L_P is an isomorphism. Let $W = R(L_A)$, define $W' = \{L_P(w) : w \in W\}$, and prove that $W' = R(L_{PA})$. □

Corollary 3. *If* $A \rightsquigarrow B$ *entirely by row operations, then* $\mathrm{rank}(A) = \mathrm{rank}(B)$.

Corollary 4. *If* $A \rightsquigarrow B$, *then* $\mathrm{rank}(A) = \mathrm{rank}(B)$.

Corollary 5. *If* $A \rightsquigarrow \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix} = D$, *then* $\mathrm{rank}(A) = r$.

Proof. Obviously $\operatorname{rank}(D) = r$, and $\operatorname{rank}(A) = \operatorname{rank}(D)$ by Corollary 4. $\square$

Corollary 6. $\operatorname{rank}(A) = \operatorname{rank}(A^t)$. *Equivalently, the row space and column space of A have the same dimension.*

Proof. $A \rightsquigarrow \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$ for some r, by the Feb 26 lecture. Then $\operatorname{rank}(A) = r$ by Corollary 5. Also

$$A^t \rightsquigarrow \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}^t = \begin{pmatrix} I_r & (O_2)^t \\ (O_1)^t & (O_3)^t \end{pmatrix}$$

by the 1st Corollary today, so $\operatorname{rank}(A^t) = r$ by Corollary 5. Hence $\operatorname{rank}(A^t) = r = \operatorname{rank}(A)$. $\square$

Recall from Feb 26 that if $A \in M_{n\times n}(\mathbb{F})$ then A can be transformed by elementary row and column operations to $D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$ where $r = \operatorname{rank}(A)$. Furthermore,

$$D = \underbrace{(E_k \cdots E_2 E_1)}_{P} A \underbrace{(E_1' E_2' \cdots E_\ell')}_{Q} = PAQ$$

where E_i, E_j' are the corresponding elementary matrices (E_i for row ops, E_j' for column ops).

Theorem (Invertible Matrix Theorem)**.** *For $A \in M_{n\times n}(\mathbb{F})$, the following are equivalent:*

(1) *A is invertible.*
(2) $\operatorname{rank}(A) = n$.
(3) *A can be written as a product of elementary matrices.*
(4) $A \rightsquigarrow I_n$.
(5) *A can be transformed by elementary row operations to I_n.*

Proof sketch. (2) $\Leftrightarrow$ (1): $\operatorname{rank}(A) = n \iff L_A$ is surjective $\iff L_A$ is an isomorphism (Jan 31).

(2) $\Leftrightarrow$ (4): This follows from the remarks before the Theorem.

(4) $\Rightarrow$ (3): If $A \rightsquigarrow I_n$, then $I_n \rightsquigarrow A$. Thus

$$A \;=\; (E_k \cdots E_2 E_1) I_n (E_1' E_2' \cdots E_\ell') \;=\; E_k \cdots E_2 E_1 E_1' E_2' \cdots E_\ell'.$$

(3) $\Rightarrow$ (5): Assume that $A = E_1 E_2 \cdots E_k$ where each E_i is elementary. Then A is invertible (it is a product of invertible matrices), and $A^{-1} = E_k^{-1} \cdots E_2^{-1} E_1^{-1}$. Thus

$$I_n \;=\; A^{-1}A \;=\; E_k^{-1} \cdots E_2^{-1} E_1^{-1} A.$$

Each E_i^{-1} is also elementary. Multiplying on the left by elementary matrices is the same as applying elementary row operations. Thus the equation $I_n = E_k^{-1} \cdots E_2^{-1} E_1^{-1} A$ shows that A can be transformed by elementary row operations to I_n. $\square$

Here is an application. Suppose A is invertible, so can be transformed by elementary row operations to I_n. Let $E_1, \ldots, E_k$ be the corresponding elementary matrices. Thus $I_n = E_k \cdots E_2 E_1 A$. Multiply both sides of this equation on the right by A^{-1} to get $A^{-1} = E_k \cdots E_2 E_1 I_n$. This proves the following:

Theorem. *If A is invertible, then the sequence of elementary row operations which transforms A to I_n also transforms I_n to A^{-1}.*

This theorem gives an easy way to find A^{-1} (when it exists).

(1) Form the $n \times 2n$ matrix $(A \,|\, I_n)$.
(2) Using row operations, transform A to I_n, but apply the operations to $(A \,|\, I_n)$.
(3) If $(A \,|\, I_n)$ is transformed to $(I_n \,|\, B)$, then $B = A^{-1}$.

The above algorithm can be applied to a square matrix A even when A is not invertible. Here is what will happen.

If A is not invertible, then it can be shown that the attempt to transform A to I_n via row operations will always lead to a row of all zeroes. Elementary operations do not change the rank of a matrix (Corollary 4 from Feb. 28). Clearly if an $n \times n$ matrix A' has a row of all zeroes, then the row space of A' has dimension at most $n - 1$. Since the dimension of the row space of A' equals the rank of A' by Corollary 6 from Feb. 28, it follows that $\operatorname{rank}(A') < n$. This proves that if elementary row operations transform A to a matrix A' having a row of all zeroes, then $\operatorname{rank}(A) < n$ and so A is not invertible.

Thus in the process of transforming $(A \,|\, I_n)$, if at any point you arrive at the situation where a row has the form $(0 \cdots 0 \,|\, * \cdots *)$, you can stop and conclude that A is not invertible.

Consider a system of m linear equations in n unknowns:

$$(S)\qquad \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & \cdots & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & \cdots & a_{2n}x_n & = & b_2 \\ & & & & & \vdots & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & \cdots & a_{mn}x_n & = & b_m \end{array}$$

We always have a field $\mathbb{F}$ in mind. The coefficients a_{ij} and right-hand sides b_i belong to $\mathbb{F}$, and we are looking for solutions $(x_1, \ldots, x_n) \in \mathbb{F}^n$.

We can write the system as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_1 \\ \vdots \\ b_m \end{pmatrix},$$

that is, as

$$AX = b$$

where $A \in M_{m\times n}(\mathbb{F})$, $b \in \mathbb{F}^m$, and $X = (x_1, \ldots, x_n)$ is an n-tuple of formal variables "ranging over $\mathbb{F}$."

- The formal matrix equation $AX = b$ is the **matrix form** of (S).
- A is the **coefficient matrix** and b is the **RHS vector**.
- $(A \,|\, b) \in M_{m\times(n+1)}(\mathbb{F})$ is called the **augmented matrix** of the system.
- The system is said to be **homogeneous** if $b = 0$, and is **nonhomogeneous** otherwise.
- If $b \neq 0$, then the system $AX = 0$ is the **homogeneous system associated to** $AX = b$.

Definition. A **solution** to $AX = b$ is a vector $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$ satisfying $Ax = b$. The **solution set** to $AX = b$ is the set

$$\mathrm{Sol}(AX{=}b) \;=\; \{x \in \mathbb{F}^n \,:\, Ax = b\}.$$

A system is **consistent** if it has at least one solution; otherwise it is **inconsistent**.

Theorem 3.15. *Let $A \in M_{m\times n}(\mathbb{F})$ and $b \in \mathbb{F}^m$.*

(1) $\mathrm{Sol}(AX{=}0) = N(A)$.
(2) *$AX = b$ is consistent iff $b \in$ the column space of A.*
(3) *If $AX = b$ is consistent, then its solution set is a translation of $N(A)$, i.e.,*

$$\mathrm{Sol}(AX{=}b) \;=\; u + N(A) \qquad \text{where } u \text{ can be any solution to } AX = b.$$

Proof. In class. □

Theorem 3.16. *Suppose $A \in M_{m\times n}(\mathbb{F})$, $b \in \mathbb{F}^m$, and the augmented matrix $(A \,|\, b)$ can be transformed via elementary row operations to a matrix $(A' \,|\, b')$. Then $\mathrm{Sol}(AX{=}b) = \mathrm{Sol}(A'X{=}b')$.*

Proof. In class. □

Definition. A matrix is **in Reduced Row Echelon Form** (or **RREF**) if all of the following hold:

(1) If a row has a nonzero entry, then its first nonzero entry is 1. (Called the **leading 1** of the row.)
(2) If a column contains a leading 1 (of some nonzero row), then all other entries in the column are 0.
(3) Lower nonzero rows have their leading 1 increasingly to the right.
(4) All-zero rows (if any) are at the bottom of the matrix.

Theorem. *For every matrix $A \in M_{m\times n}(\mathbb{F})$ there exists a matrix R in RREF such that $A \rightsquigarrow R$ via row operations.*

Proof sketch. Given in class. $\square$

It is easy to determine the rank of a matrix in RREF.

Proposition. *If R is in RREF, then* $\operatorname{rank}(R) =$ *the number of leading 1s.*

Proof. Suppose the first k rows of R have leading 1s and the rest of the rows are zero. The columns containing the leading 1s are $e_1, \dots, e_k$ and they clearly form a basis for the columns space of R. $\square$

We can apply these results to augmented matrices of linear systems. If $(A \,|\, b) \rightsquigarrow (R \,|\, s)$ by row operations then $\operatorname{Sol}(AX = b) = \operatorname{Sol}(RX = s)$. So it suffices to describe $\operatorname{Sol}(RX = s)$ when $(R \,|\, s)$ is in RREF. I will illustrate the method by an example. Suppose

$$(R \,|\, s) = \left(\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & -3 & s_1 \\ 0 & 1 & -1 & 0 & 4 & s_2 \\ 0 & 0 & 0 & 1 & -2 & s_3 \\ 0 & 0 & 0 & 0 & 0 & s_4 \end{array}\right).$$

- Obviously column space$(R) = \operatorname{span}\{e_1, e_2, e_3\}$ so $RX = s$ is consistent iff $s_4 = 0$.
- Assuming $s_4 = 0$, write the equations of the system corresponding to $RX = s$:

$$\begin{cases} x_1 + 2x_3 - 3x_5 = s_1 \\ x_2 - x_3 + 4x_5 = s_2 \\ x_4 - 2x_5 = s_3 \\ 0 = 0 \end{cases}$$

- The variables corresponding to leading 1s are said to be *dependent*; the other variables are *free*. In this example, x_3 and x_5 are free.
- Rewrite the system by expressing each variable in terms of the free variables:

$$\begin{cases} x_1 = s_1 + -2x_3 + 3x_5 \\ x_2 = s_2 + x_3 - 4x_5 \\ x_3 = x_3 \\ x_4 = s_3 + 2x_5 \\ x_5 = x_5 \end{cases}$$

- Now rewrite these last equations in vector form:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \underset{\substack{\uparrow \\ u}}{\begin{pmatrix} s_1 \\ s_2 \\ 0 \\ s_3 \\ 0 \end{pmatrix}} + x_3 \underset{\substack{\uparrow \\ v_1}}{\begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}} + x_5 \underset{\substack{\uparrow \\ v_2}}{\begin{pmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{pmatrix}}.$$

- Note that x_3, x_5 can be any values in $\mathbb{F}$. Thus

$$\operatorname{Sol}(AX = b) = u + \operatorname{span}(\{v_1, v_2\}).$$

- It is easy to check that u is a solution to $RX = s$ and so is also a solution to $AX = b$. The same analysis with s replaced by 0 shows that $\operatorname{span}\{v_1, v_2\} = N(R) = N(A)$. In this example $\operatorname{rank}(A) = \operatorname{rank}(R) = 3$ so $\operatorname{nullity}(A) = 5 - 3 = 2$. Thus $\{v_1, v_2\}$ is a basis for $N(A)$.

The same arguments work generally. If $A \in M_{m\times n}(\mathbb{F})$, $b \in \mathbb{F}^m$, and $(A \,|\, b) \overset{row}{\rightsquigarrow} (R \,|\, s)$ in RREF, $RX = s$ is consistent (i.e., $(R \,|\, s)$ has no row of the form $(0 \cdots 0 \,|\, 1)$), and R has r columns containing a leading 1 and hence $n - r$ free variables, then $\operatorname{nullity}(A) = n - r$ and the equations for $RX = s$, when translated to expressions for each variable in terms of the free variables and then written in vector form, lead to a description of

$$\operatorname{Sol}(AX{=}b) = u + \operatorname{span}(\{v_1, \dots, v_{n-r}\})$$

where $\operatorname{span}(\{v_1, \dots, v_{n-r}\}) = N(A)$. Hence u is a solution to $AX = b$ and $\{v_1, \dots, v_{n-r}\}$ is a basis for $N(A)$.

Theorem. *For each $A \in M_{m\times n}(\mathbb{F})$, there is a unique matrix R in RREF such that $A \overset{row}{\rightsquigarrow} R$.*

Proof sketch. Given A, there is at least one such R (by today's first theorem). It suffices to show that the entries of R are determined by A.

By A4Q5(b), the columns of R that contain leading 1s are exactly the columns with index j such that $\operatorname{Col}_j(A) \notin \operatorname{span}\{\operatorname{Col}_1(A), \dots, \operatorname{Col}_{j-1}(A)\}$. In this way A determines the indices $j_1, \dots, j_r$ of columns of R containing leading 1s. By definition of RREF, $\operatorname{Col}_{j_t}(R) = e_t$ for $t = 1, \dots, r$.

Let $j \in \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$. I.e., j is the index of a column of R which does not contain a leading 1. By definition of RREF, if $j < j_1$ then $\operatorname{Col}_j(R) = 0$. Otherwise, let $t \in \{1, \dots, r\}$ be the largest such that $j > j_t$. Then by definition of RREF,

$$\operatorname{Col}_j(R) = \begin{pmatrix} c_1 \\ \vdots \\ c_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = c_1 \operatorname{Col}_{j_1}(R) + \cdots + c_t \operatorname{Col}_{j_t}(R).$$

Apply A4Q5(a) to get

$$(*) \qquad \operatorname{Col}_j(A) = c_1 \operatorname{Col}_{j_1}(A) + \cdots + c_t \operatorname{Col}_{j_t}(A).$$

We know that $\operatorname{Col}_{j_1}(R), \dots, \operatorname{Col}_{j_t}(R)$ are linearly independent; thus by A4Q5(a) again, $\operatorname{Col}_{j_1}(A), \dots, \operatorname{Col}_{j_t}(A)$ are linearly independent. So the equation $(*)$ uniquely determines $c_1, \dots, c_t$. □

To every square matrix $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ there is an associated scalar $\det(A) \in \mathbb{F}$, called the **determinant** of A. In the next few lectures I will give a definition of $\det(A)$ and discuss/prove a number of properties of $\det(-)$ as a function $M_{n\times n}(\mathbb{F}) \to \mathbb{F}$. I'll start with the 2×2 case.

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2\times 2}(\mathbb{F})$, then $\det(A) = ad - bc$.

Facts (2×2 case)

(1) A is invertible (i.e., $\operatorname{rank}(A) = 2$) iff $\det(A) \neq 0$.

(2) If A is invertible, then $A^{-1} = \dfrac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

(3) If $A, B \in M_{2\times 2}(\mathbb{F})$, then $\det(AB) = \det(A) \cdot \det(B)$.

The definition of $\det(A)$ in higher dimensions will have these (and other) properties.

First, to any square matrix we assign $(-1)^{i+j}$ to the (i,j) position. This creates a "checkerboard" sign pattern

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

Definition. Suppose $A \in M_{n\times n}(\mathbb{F})$ and $1 \leq i, j \leq n$.

(1) $\widetilde{A}_{ij}$ denotes the $n-1 \times n-1$ matrix obtained by deleting the i-th row and the j-th column from A.

(2) $\widetilde{A}_{ij}$ is called the (i,j)-**submatrix** of A.

Once determinants have been defined:

(3) $\det(\widetilde{A}_{ij})$ will be called the (i,j) **minor** of A.

(4) $(-1)^{i+j}\det(\widetilde{A}_{ij})$ will be called the (i,j) **cofactor** of A.

Recursive definition of det.

(1) If $A = (a) \in M_{1\times 1}(\mathbb{F})$, then $\det(A) = a$.

(2) If $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in M_{n\times n}(\mathbb{F})$ with $n > 1$, then

$$\begin{aligned} \det(A) &= a_{11} \cdot \det(\widetilde{A}_{11}) - a_{21} \cdot \det(\widetilde{A}_{21}) + a_{31} \cdot \det(\widetilde{A}_{31}) - \cdots \\ &= \sum_{i=1}^{n} a_{i1} \cdot \underbrace{(-1)^{i+1}\det(\widetilde{A}_{i1})}_{(i,1)\, cofactor\, of\, A}. \end{aligned}$$

The recursive definition in (2) is called **expansion by minors (or cofactors) on the first column**. To prove facts about $\det(-)$, we deal directly with its recursive definition.

Lemma 4.0. *If $A \in M_{n\times n}$ is **upper-triangular** (i.e., $a_{ij} = 0$ whenever $i > j$), then* $\det(A) = \prod_{i=1}^{n} a_{ii}$.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2,n-1} & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3,n-1} & a_{3n} \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

Proof. By induction on n (in class). □

Corollary 4.1. $\det(I_n) = 1$.

Theorem 4.2. *If $A \in M_{n\times n}(\mathbb{F})$ and A has a row of zeros, then* $\det(A) = 0$.

Proof. By induction on n (in class). □

Theorem 4.3. *If $A \in M_{n\times n}(\mathbb{F})$ and A has a column of zeros, then* $\det(A) = 0$.

Proof. By induction on n (in class). □

Theorem 4.5 (adj)**.** *If $A \in M_{n\times n}(\mathbb{F})$ and A has two adjacent rows that are equal, then* $\det(A) = 0$.

Proof sketch. By induction. In the inductive step, suppose $\mathrm{Row}_{i_0}(A) = \mathrm{Row}_{i_0+1}(A) = (r_1\ r_2\ \cdots\ r_n)$. Then

(1) For all $i \neq i_0, i_0{+}1$, $\widetilde{A}_{i1}$ has two equal adjacent rows so $\det(\widetilde{A}_{i1}) = 0$ by induction.
(2) $\widetilde{A}_{i_0,1} = \widetilde{A}_{i_0+1,1}$, and their determinants appear (in the definition of $\det A$) with the same coefficient (r_1) and opposite sign.

Hence everything is 0 or cancels and $\det A = 0$. □

Theorem 4.6. $\det$ *is "linear in each row." That is, if we fix n, i_0, and $u_1, \ldots, u_{i_0-1}, u_{i_0+1}, \ldots, u_n \in \mathbb{F}^n$, then for all $r, s \in \mathbb{F}^n$ and all $a \in \mathbb{F}$,*

$$\det\begin{pmatrix} \text{---}\ u_1\ \text{---} \\ \vdots \\ \text{---}\ r+s\ \text{---} \\ \vdots \\ \text{---}\ u_n\ \text{---} \end{pmatrix} = \det\begin{pmatrix} \text{---}\ u_1\ \text{---} \\ \vdots \\ \text{---}\ r\ \text{---} \\ \vdots \\ \text{---}\ u_n\ \text{---} \end{pmatrix} + \det\begin{pmatrix} \text{---}\ u_1\ \text{---} \\ \vdots \\ \text{---}\ s\ \text{---} \\ \vdots \\ \text{---}\ u_n\ \text{---} \end{pmatrix}$$

where r, s and $r+s$ were inserted in row i_0; and similarly,

$$\det\begin{pmatrix} \text{---}\ u_1\ \text{---} \\ \vdots \\ \text{---}\ ar\ \text{---} \\ \vdots \\ \text{---}\ u_n\ \text{---} \end{pmatrix} = a \cdot \det\begin{pmatrix} \text{---}\ u_1\ \text{---} \\ \vdots \\ \text{---}\ r\ \text{---} \\ \vdots \\ \text{---}\ u_n\ \text{---} \end{pmatrix}.$$

Proof sketch by example. Consider the first claim when $n = 4$ and $i_0 = 3$. Write

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ r_1 & r_2 & r_3 & r_4 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \qquad B = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ s_1 & s_2 & s_3 & s_4 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \qquad C = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ r_1{+}s_1 & r_2{+}s_2 & r_3{+}s_3 & r_4{+}s_4 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

Note that:

(1) For each $j = 1, 2, 4$ the inductive hypothesis applies to give $\det(\widetilde{C}_{j1}) = \det(\widetilde{A}_{j1}) + \det(\widetilde{B}_{j1})$.
(2) $\widetilde{C}_{31} = cof A31 = \widetilde{B}_{31}$.

Putting these facts together, we get

$$\begin{aligned} \det(C) &= a_{11} \cdot \det(\widetilde{C}_{11}) - a_{21} \cdot \det(\widetilde{C}_{21}) + (r_1 + s_1) \cdot \det(\widetilde{C}_{31}) - a_{41} \cdot \det(\widetilde{C}_{41}) \\ &= \cdots \\ &= \det(A) + \det(B). \end{aligned}$$

□

Theorem 4.7 (adj)**.** *If $A \in M_{n\times n}(\mathbb{F})$ and $A \xrightarrow{R_i \leftarrow R_i + cR_j} B$ where $j = i \pm 1$, then* $\det(B) = \det(A)$.

Proof. Let $\mathrm{Row}_i(A) = r$, $\mathrm{Row}_j(A) = s$, and $\mathrm{Row}_t(A) = u_t$ for $t \neq i, j$. Assume $j = i+1$. Thus

$$A = \begin{pmatrix} & \vdots & \\ \text{———} & r & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} & \vdots & \\ \text{———} & r + cs & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix}$$

so by "linearity in row i,"

$$\det(B) = \det \begin{pmatrix} & \vdots & \\ \text{———} & r + cs & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ \text{———} & r & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} + c \cdot \det \begin{pmatrix} & \vdots & \\ \text{———} & s & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix}.$$

The first determinant on the right side is $\det(A)$. The second determinant on the right side is 0 because the matrix has two equal adjacent rows (using Theorem 4.5(adj) here). $\square$

Theorem 4.8 (adj)**.** *If* $A \in M_{n\times n}(\mathbb{F})$ *and* $A \overset{R_i \leftrightarrows R_{i+1}}{\longrightarrow} B$, *then* $\det(B) = -\det(A)$.

Proof. This can be deduced from Theorem 4.7(adj) and linearity in rows i and $i+1$, as follows.

$$\begin{aligned} \det(B) &= \det \begin{pmatrix} & \vdots & \\ \text{———} & s & \text{———} \\ \text{———} & r & \text{———} \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ \text{———} & s - r & \text{———} \\ \text{———} & r & \text{———} \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ \text{———} & s - r & \text{———} \\ \text{———} & r + (s-r) & \text{———} \\ & \vdots & \end{pmatrix} \\ &= \det \begin{pmatrix} & \vdots & \\ \text{———} & s - r & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} = \det \begin{pmatrix} & \vdots & \\ \text{———} & (-r) & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} = -\det \begin{pmatrix} & \vdots & \\ \text{———} & r & \text{———} \\ \text{———} & s & \text{———} \\ & \vdots & \end{pmatrix} = -\det(A). \end{aligned}$$

$\square$

Theorem 4.5 (gen)**.** *If A has two equal rows (not necessarily adjacent), then* $\det(A) = 0$.

Proof. By a sequence of adjacent row switches, we can transform A to a matrix A' with two equal adjacent rows. Then $\det(A) = \pm\det(A') = 0$ by Theorems 4.8(adj) and 4.5(adj). $\square$

Theorem 4.7 (gen)**.** *If* $A \in M_{n\times n}(\mathbb{F})$ *and* $A \overset{R_i \leftarrow R_i + cR_j}{\longrightarrow} B$, *then* $\det(B) = \det(A)$.

Proof. Just like the proof of Theorem 4.7(adj), using Theorem 4.5(gen) instead of Theorem 4.5(adj). $\square$

Theorem 4.8 (gen)**.** *If* $A \overset{R_i \leftrightarrows R_j}{\longrightarrow} B$ *then* $\det(B) = -\det(A)$.

Proof. Just like the proof of Theorem 4.8(adj), using Theorem 4.7 (gen) instead of Theorem 4.7(adj). $\square$

Corollary 4.9. *If* $A \overset{R_i \leftarrow cR_i}{\longrightarrow} B$ *then* $\det(B) = c \cdot \det(A)$.

Proof. By linearity of $\det(-)$ in row i. $\square$

Theorem 4.7 (gen). *If* $A \in M_{n\times n}(\mathbb{F})$ *and* $A \xrightarrow{R_i \leftarrow R_i + cR_j} B$, *then* $\det(B) = \det(A)$.

Corollary 4.8 (gen). *If* $A \xrightarrow{R_i \leftrightarrows R_j} B$ *then* $\det(B) = -\det(A)$.

Corollary 4.9. *If* $A \xrightarrow{R_i \leftarrow cR_i} B$ *then* $\det(B) = c \cdot \det(A)$.

Hence we completely understand the effect of elementary row operations on the value of $\det(-)$. Recall that elementary matrices can be obtained by applying row elementary operations to I_n. Thus:

(1) If E is an elementary matrix of the first kind ($R_i \leftrightarrows R_j$), then $\det(E) = -\det(I_n) = -1$.
(2) If E is an elementary matrix of the second kind ($R_i \leftarrow cR_i$), then $\det(E) = c \cdot \det(I_n) = c$.
(3) If E is an elementary matrix of the third kind ($R_i \leftarrow R_i + cR_j$) then $\det(E) = \det(I_n) = 1$.

Note that $\det(E) \neq 0$ for every elementary matrix. Also note that $\det(E^t) = \det(E)$ because E^t is an elementary matrix of the same type as E.

We pause to note that these facts give us an efficient way to calculate the determinant of a matrix: transform the matrix to upper-triangular form using elementary row operations of types 1 and 3 only.

Example. To find $\det \begin{pmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 3 & -1 & 1 \end{pmatrix}$, do

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -2 & -3 & -5 \\ 3 & -1 & 1 \end{pmatrix} \xrightarrow{(1)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 3 & -1 & 1 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 0 & -\frac{11}{2} & -\frac{13}{2} \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} -2 & -3 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 10 \end{pmatrix} = B.$$

B is upper-triangular, so $\det(B) = (-2)\cdot 1 \cdot 10 = -20$. We swapped rows once, so $\det(A) = -\det(B) = 20$.

In general, count the number of times two rows are swapped, and change the sign if the number is odd.

(Back to the general theory.)

Theorem 4.10. *If* $A, E \in M_{n\times n}(\mathbb{F})$ *with* E *elementary, then* $\det(EA) = \det(E)\det(A)$.

Proof. Let $\mathcal{O}$ be the elementary row operation corresponding to E. Then $A \xrightarrow{\mathcal{O}} EA$. Theorem 4.7(gen) and Corollaries 4.8 and 4.9 relate $\det(EA)$ to $\det(A)$ via a constant factor of -1, c, or 1 (depending on the type of operation). Since this constant factor is equal to $\det(E)$ (see remarks above), we get $\det(EA) = \det(E)\det(A)$. $\square$

Let's extend this last result. Suppose $A, B \in M_{n\times n}(\mathbb{F})$ and

$$B = E_1 E_2 \cdots E_k A$$

where $E_1, \ldots, E_k$ are elementary matrices. By applying Theorem 4.10 repeatedly, we get

$$\begin{aligned} \det(B) &= \det(E_1(E_2E_3\cdots E_kA)) \\ &= \det(E_1)\det(E_2E_3\cdots E_kA) \qquad \text{by Theorem 4.10} \\ &= \det(E_1)\det(E_2)\det(E_3\cdots E_kA) \qquad \text{by Theorem 4.10} \\ &\;\;\vdots \\ \text{and so} \qquad \det(E_1E_2\cdots E_kA) &= \det(E_1)\det(E_2)\cdots\det(E_k)\det(A). \qquad (*) \end{aligned}$$

I'm going to call this result $(*)$ and draw some consequences from it.

(1) If we set $A = I_n$ in $(*)$, we get

$$\det(E_1 E_2 \cdots E_k) = \det(E_1)\det(E_2)\cdots\det(E_k). \qquad (**)$$

(2) Since every invertible matrix can be written as a product of elementary matrices, and the determinant of any elementary matrix is nonzero, this proves that A invertible $\implies \det(A) \neq 0$.

(3) Suppose A is not invertible. Then $\operatorname{rank}(A) < n$. A can be transformed by elementary row operations to some matrix R in RREF. Necessarily R has a row of zeroes. Thus $\det(R) = 0$ by Theorem 4.3. Furthermore, R can be transformed to A by elementary row operations, so $A = E_1 E_2 \cdots E_k R$ for some elementary matrices $E_1, \ldots, E_k$. It follows from $(*)$ that

$$\det(A) = \det(E_1)\det(E_2)\cdots\det(E_k)\det(R) = 0 \qquad \text{since } \det(R) = 0.$$

Items (2) and (3) prove:

Theorem 4.11. *If $A \in M_{n\times n}(\mathbb{F})$, then A is invertible iff $\det(A) \neq 0$.*

Now we can prove

Theorem 4.12. *For all $A, B \in M_{n\times n}(\mathbb{F})$, $\det(AB) = \det(A)\det(B)$.*

Proof. Case 1: A is invertible. Then A can be written as a product of elementary matrices:

$$A = E_k \cdots E_2 E_1.$$

Thus

$$\begin{aligned} \det(AB) &= \det(E_k \cdots E_2 E_1 B) \\ &= \det(E_k)\cdots\det(E_2)\det(E_1)\det(B) && \text{by } (*) \\ &= \det(E_k \cdots E_2 E_1)\det(B) && \text{by } (**) \\ &= \det(A)\det(B). \end{aligned}$$

Case 2: A is not invertible. Then AB is also not invertible (Feb. 14), so

$$\det(A)\det(B) = 0 \cdot \det(B) = 0 = \det(AB)$$

by Theorem 4.11. □

Corollary 4.13. *If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.*

Proof. By Theorem 4.12, $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1$. □

Corollary 4.14. $\det(A^t) = \det(A)$.

Proof. We already know this for elementary matrices. Now consider cases.

Case 1: $\operatorname{rank}(A) < n$. Then $\operatorname{rank}(A^t) < n$ as well so $\det(A) = 0 = \det(A^t)$ by Theorem 4.11.

Case 2: $\operatorname{rank}(A) = n$. Then we can write $A = E_1 E_2 \cdots E_k$ with each E_i elementary. By Theorem 4.12, $\det(A) = \det(E_1)\det(E_2)\cdots\det(E_k)$. We also have $A^t = (E_k)^t \cdots (E_2)^t (E_1)^t$, which is a product of elementary matrices, so again by Theorem 4.12,

$$\begin{aligned} \det(A^t) &= \det((E_k)^t)\cdots\det((E_2)^t)\det((E_1)^t) \\ &= \det(E_k)\cdots\det(E_2)\det(E_1) && \text{as } \det(E^t) = \det(E) \text{ for elementary } E \\ &= \det(A). \end{aligned}$$

□

Welcome back! Recall from way back when:

Corollary 4.14. $\det(A^t) = \det(A)$.

This corollary implies that if we have proved a determinant result about rows, we can deduce the same result for columns. For example:

Corollary 4.15 ("Column version" of Corollary 4.8(gen))**.** *If* $A \xrightarrow{C_i \leftrightarrows C_j} B$ *then* $\det(B) = -\det(A)$.

Corollary 4.16 ("Column version" of Theorem 4.6)**.** $\det(-)$ *is "linear in each column."*

Suppose $A = \begin{pmatrix} \text{---}\ r_1\ \text{---} \\ \text{---}\ r_2\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix}$ and $B = \begin{pmatrix} \text{---}\ r_2\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_1\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix}$. Thus B is obtained from A by "cyclically permuting" the first 4 rows. How are $\det(A)$ and $\det(B)$ related? We can simulate the cyclic permutation by a sequence of row swaps:

$$A = \begin{pmatrix} \text{---}\ r_1\ \text{---} \\ \text{---}\ r_2\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix} \xrightarrow{R_1 \leftrightarrows R_2} \begin{pmatrix} \text{---}\ r_2\ \text{---} \\ \text{---}\ r_1\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix} \xrightarrow{R_2 \leftrightarrows R_3} \begin{pmatrix} \text{---}\ r_2\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_1\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix} \xrightarrow{R_3 \leftrightarrows R_4} \begin{pmatrix} \text{---}\ r_2\ \text{---} \\ \text{---}\ r_3\ \text{---} \\ \text{---}\ r_4\ \text{---} \\ \text{---}\ r_1\ \text{---} \\ \text{---}\ r_5\ \text{---} \\ \vdots \end{pmatrix} = B.$$

Three row-swaps were used, so $\det(B) = -\det(A)$. In general,

Lemma 4.17. *If* $A \in M_{n\times n}(\mathbb{F})$ *and* B *is obtained from* A *by cyclically permuting* k *consecutive rows, then* $\det(B) = (-1)^{k-1}\det(A)$.

Note: a similar result holds for cyclically permuting k consecutive columns of A (via $\det(A) = \det(A^t)$).

We are now ready to tackle a technically difficult result: the "Lemma before Theorem 4.4 in the textbook."

Lemma 4.18. *Suppose* $A \in M_{n\times n}(\mathbb{F})$ *and for some* $1 \le i, j \le n$, *row* i *of* A *is* $e_j = (0, \dots, 0, 1, 0, \dots, 0)$. *Then* $\det(A) = (-1)^{i+j}\det(\widetilde{A}_{ij})$.

Proof. Case 1: $j = 1$. Then A has the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & & & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \leftarrow i$$

When we calculate $\det(A)$ using the recursive definition, the minors $\det(\widetilde{A}_{t1})$ will all have a row of zeros (so will equal 0), except for the minor at row i. Thus $\det(A) = (-1)^{i+1}\det(\widetilde{A}_{i1})$ in Case 1.

Case 2: $j > 1$. Let B be the matrix obtained from A by cyclically shifting the first j *columns*. Thus

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,j-1} & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,j-1} & a_{2j} & \cdots & a_{2n} \\ \vdots & & & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,j-1} & a_{nj} & \cdots & a_{nn} \end{pmatrix}, \qquad B = \begin{pmatrix} a_{1j} & a_{11} & a_{12} & \cdots & a_{1,j-1} & \cdots & a_{1n} \\ a_{2j} & a_{21} & a_{22} & \cdots & a_{2,j-1} & \cdots & a_{2n} \\ \vdots & & & & & & \vdots \\ 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ a_{nj} & a_{n1} & a_{n2} & \cdots & a_{n,j-1} & \cdots & a_{nn} \end{pmatrix}.$$

We have:

(1) $\det(A) = (-1)^{j-1}\det(B)$ by the "column" version of Lemma 4.17.
(2) $\det(B) = (-1)^{i+1}\det(\widetilde{B}_{i1})$ by the proof of Case 1.
(3) $\widetilde{B}_{i1} = \widetilde{A}_{ij}$ (easily seen).

Hence $\det(A) = (-1)^{j-1}(-1)^{i+1}\det(\widetilde{B}_{i1}) = (-1)^{i+j}\det(\widetilde{A}_{ij})$ in Case 2. □

Now we are ready for the biggest theorem.

Theorem 4.19. $\det(-)$ *can be evaluated by expansion by cofactors on any row, or any column.*
That is, if $A \in M_{n\times n}(\mathbb{F})$*, then*

- *For any* $i = 1, \dots, n$*,* $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(\widetilde{A}_{ij})$. (Expansion on row i)
- *For any* $j = 1, \dots, n$*,* $\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(\widetilde{A}_{ij})$. (Expansion on column j)

Proof. I'll prove the claim for expansion on a row. Fix i and consider a matrix

$$A = \begin{pmatrix} \text{——} \; r_1 \; \text{——} \\ \vdots \\ \text{——} \; r_i \; \text{——} \\ \vdots \\ \text{——} \; r_n \; \text{——} \end{pmatrix}.$$

Write $r_i = (a_{i1}, a_{i2}, \dots, a_{in}) = a_{i1}e_1 + a_{i2}e_2 + \cdots + a_{in}e_n$. Then

$$\begin{aligned} \det(A) &= \det \begin{pmatrix} \text{——} \; r_1 \; \text{——} \\ \vdots \\ a_{i1}e_1 + \cdots + a_{in}e_n \\ \vdots \\ \text{——} \; r_n \; \text{——} \end{pmatrix} = \sum_{j=1}^{n} a_{ij} \det \begin{pmatrix} \text{——} \; r_1 \; \text{——} \\ \vdots \\ \text{——} \; e_j \; \text{——} \\ \vdots \\ \text{——} \; r_n \; \text{——} \end{pmatrix} && \text{by linearity in row } i \\ &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(\widetilde{A}_{ij}) && \text{by Lemma 4.18.} \end{aligned}$$

Look at that! This is the formula for the expansion by cofactors on row i.
The claim for columns can be obtained from the claim from rows by taking transposes. □

Theorem 4.19 can make the calculation of some determinants quite easy.

Example. Let $A = \begin{pmatrix} 2 & 0 & -1 & 3 \\ 3 & -2 & 4 & 0 \\ 6 & 1 & 3 & 0 \\ 7 & 2 & 0 & 5 \end{pmatrix}$. By Theorem 4.19, we can calculate $\det(A)$ by cofactor expansion along *any* row or column. The 4th column conveniently has 2 zeros, so let's use it. Recalling that the $\pm$ pattern in column 4 starts with $-$, we get

$$\det(A) = (-3)\det\underbrace{\begin{pmatrix} 3 & -2 & 4 \\ 6 & 1 & 3 \\ 7 & 2 & 0 \end{pmatrix}}_{=:B} + (0)(\text{something}) - (0)(\text{something}) + (5)\det\underbrace{\begin{pmatrix} 2 & 0 & -1 \\ 3 & -2 & 4 \\ 6 & 1 & 3 \end{pmatrix}}_{=:C}.$$

To compute $\det(B)$, we might note that B has a zero in row 3, so we decide to use cofactor expansion on the 3rd row:

$$\begin{aligned} \det(B) &= (7)\det\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix} - (2)\det\begin{pmatrix} 3 & 4 \\ 6 & 3 \end{pmatrix} + (0)(\text{something}) \\ &= 7(-6-4) - (2)(9-24) \\ &= -40. \end{aligned}$$

Similarly, to compute $\det(C)$ we might note that C has a zero in row 1, so we decide to use cofactor expansion on the 1st row:

$$\begin{aligned} \det(C) &= (2)\det\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix} - (0)(\text{something}) + (-1)\det\begin{pmatrix} 3 & -2 \\ 6 & 1 \end{pmatrix} \\ &= 2(-6-4) + (-1)(3+12) \\ &= -35. \end{aligned}$$

Then

$$\det(A) = (-3)(-40) + (5)(-35) = -55.$$

Announcement

- Before reading these notes, read the Tutorial on "Permutations and Determinants"

Definition. Let $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. An **eigenvector** of A is any nonzero vector $v \in \mathbb{F}^n$ satisfying $Av \in \operatorname{span}(v)$. The (unique) scalar $\lambda \in \mathbb{F}$ satisfying $Av = \lambda v$ is the **eigenvalue** of A corresponding to v.

Definition. Suppose $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. If $\lambda \in \mathbb{F}$ is an eigenvalue of A, the set

$$\begin{aligned} E_\lambda &= \{v \in \mathbb{F}^n : Av = \lambda v\} \\ &= \{\text{eigenvectors of } A \text{ corresponding to } \lambda\} \cup \{0\}. \end{aligned}$$

is called the **eigenspace** of A corresponding to λ.

Example. Let $A = \begin{pmatrix} 3/2 & -1 \\ 1/2 & 0 \end{pmatrix}$. If $v_1 = (1,1)$ and $v_2 = (2,1)$, then $L_A(v_1) = \frac{1}{2}v_1$ and $L_A(v_2) = v_2$. Hence v_1 and v_2 are eigenvectors of A corresponding to the eigenvalues $\frac{1}{2}$ and 1. We'll see later that these are the only eigenvalues of A. Calculations show

$$\begin{aligned} E_{\frac{1}{2}} &= \{v \in \mathbb{R}^2 : Av = \tfrac{1}{2}v\} = \{(a,a) : a \in \mathbb{R}\} = \operatorname{span}(\{v_1\}) \\ E_1 &= \{v \in \mathbb{R}^2 : Av = v\} = \{(2a,a) : a \in \mathbb{R}\} = \operatorname{span}(\{v_2\}). \end{aligned}$$

Example. Let $B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. We saw on Jan. 27 that L_B rotates vectors in $\mathbb{R}^2$ 90° counter-clockwise. Hence if $v \in \mathbb{R}^2$ and $v \neq 0$, then $L_B(v) \notin \operatorname{span}(v)$. Thus B has no eigenvectors, and hence no eigenvalues.

There is a nice way to characterize the eigenvalues of a matrix. Suppose $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff \exists v \in \mathbb{F}^n,\ v \neq 0, \text{ such that } Av = \lambda v \\ &\iff \exists v \in \mathbb{F}^n,\ v \neq 0, \text{ such that } Av - \lambda v = 0 \\ &\iff \exists v \in \mathbb{F}^n,\ v \neq 0, \text{ such that } (A - \lambda I_n)v = 0 \\ &\iff N(A - \lambda I_n) \neq \{0\} \\ &\iff \operatorname{nullity}(A - \lambda I_n) > 0 \\ &\iff \operatorname{rank}(A - \lambda I_n) < n \\ &\iff \det(A - \lambda I_n) = 0. \end{aligned}$$

We have proved:

Theorem 5.1. *Let $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$. λ is an eigenvalue of A iff $\det(A - \lambda I_n) = 0$.*

Note also that if λ is an eigenvalue of A, then $E_\lambda = N(A - \lambda I_n)$. This proves that E_λ is a subspace of $\mathbb{F}^n$. We also get $\dim(E_\lambda) = \operatorname{nullity}(A - \lambda I_n) > 0$.

If we view λ as a variable ranging over $\mathbb{F}$, then the expression $\det(A - \lambda I_n)$ defines a function $\mathbb{F} \to \mathbb{F}$ (sending $\lambda \mapsto \det(A - \lambda I_n)$). This function will turn out to be a polynomial function in λ. To make this a formal polynomial, we replace λ with an indeterminate (formal variable), t, and consider the matrix

$$A - tI_n = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} - \begin{pmatrix} t & 0 & \cdots & 0 \\ 0 & t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & t \end{pmatrix} = \begin{pmatrix} a_{11}-t & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-t & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-t \end{pmatrix}.$$

Definition. Let $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. The **characteristic polynomial** of A is the formal expression $\det(A-tI_n)$. It is denoted $p_A(t)$.

Example. If $A = \begin{pmatrix} 3/2 & -1 \\ 1/2 & 0 \end{pmatrix}$, then

$$\begin{aligned} p_A(t) &= \det\begin{pmatrix} 3/2-t & -1 \\ 1/2 & -t \end{pmatrix} = (\tfrac{3}{2}-t)(-t) - \tfrac{1}{2}(-1) \\ &= t^2 - \tfrac{3}{2}t + \tfrac{1}{2} = (t-\tfrac{1}{2})(t-1). \end{aligned}$$

The roots of $p_A(t)$ are $1/2$ and 1, which are the eigenvalues of A.

For any $A \in \mathsf{M}_{n\times n}(\mathbb{F})$, the eigenvalues of A are the scalars $\lambda \in \mathbb{F}$ which make $\det(A-\lambda I_n) = 0$, i.e., $p_A(\lambda) = 0$. Hence $\boxed{\text{the eigenvalues of } A \text{ are the roots of } p_A(t) \textit{ which belong to the scalar field } \mathbb{F}.}$

Definition. Given a square matrix $A \in M_{n\times n}(\mathbb{F})$, the *trace* of A, denoted $\operatorname{tr}(A)$, is the sum of the diagonal entries of A.

Theorem 5.2. *Let $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. Then $p_A(t)$ is a polynomial in $\mathbb{F}[t]$ of degree n. Moreover,*

(1) *The leading coefficient of $p_A(t)$ is $(-1)^n$.*
(2) *The coefficient of t^{n-1} in $p_A(t)$ is $(-1)^{n-1}\operatorname{tr}(A)$.*
(3) *The constant coefficient is $\det(A)$.*

That is,

$$p_A(t) = (-1)^n(t^n - \operatorname{tr}(A)\cdot t^{n-1}) + \cdots + \det(A).$$

Proof sketch. In the tutorial notes on "Permutations and Determinants," you learned that the determinant of a square matrix can be written as the alternating sum of all possible products consisting of one entry from each row and each column of the matrix. (This is the *complete expansion* of the determinant.) Thus the complete expansion of $\det(A-tI_n)$ is an alternating sum of products of n entries from $A-tI_n$, one from each row and each column. Each entry of $A-tI_n$ is a polynomial in $\mathbb{F}[t]$ of degree ≤ 1. Hence each product of n entries is a polynomial in $\mathbb{F}[t]$ of degree $\le n$. As $p_A(t)$ is an alternating sum of such entries, $p_A(t)$ is also a polynomial in $\mathbb{F}[t]$ of degree $\le n$. It remains to prove that the coefficients of t^n and t^{n-1} and the constant coefficient are as claimed in the theorem.

The only contributions of t to $p_A(t)$ come from the diagonal entries of $A-tI_n$. A product in the complete expansion either has all the diagonal entries, or at most $n-2$ of them. Hence the t^n term and the t^{n-1} term of $p_A(t)$ come entirely from

$$\begin{aligned} (a_{11}-t)(a_{22}-t)\cdots(a_{nn}-t) &= (-t)^n + (-t)^{n-1}(a_{11}+\cdots+a_{nn}) + (\text{lower degree terms}) \\ &= (-1)^n\cdot t^n + (-1)^{n-1}\operatorname{tr}(A)\cdot t^{n-1} + (\text{lower degree terms}). \end{aligned}$$

This proves (1) and (2). (3) follows by setting $t=0$ in the definition of $p_A(t)$. $\square$

Corollary. *If $A \in \mathsf{M}_{n\times n}(\mathbb{F})$, then A has at most n eigenvalues.*

Proof. A polynomial of degree n has at most n roots in any field. $\square$

Definition. Suppose $A, B \in \mathsf{M}_{n\times n}(\mathbb{F})$. We say that B is **similar to** A (over $\mathbb{F}$) if there exists an invertible $Q \in \mathsf{M}_{n\times n}(\mathbb{F})$ such that $B = Q^{-1}AQ$.

"Exercise 12." If B is similar to A, then $p_B(t) = p_A(t)$.

Proof. We can write $tI_n = tQ^{-1}I_nQ = Q^{-1}(tI_n)Q$. Hence

$$B - tI_n = Q^{-1}AQ - Q^{-1}(tI_n)Q = Q^{-1}(A - tI_n)Q.$$

So

$$\begin{aligned}
p_B(t) = \det(B - tI_n) &= \det(Q^{-1}(A - tI_n)Q)) \\
&= \det(Q^{-1})\det(A - tI_n)\det(Q) && \text{as } \det(AB) = \det(A)\det(B) \\
&= \det(A - tI_n)\det(Q^{-1})\det(Q) && \text{as multiplication in } \mathbb{F} \text{ is commutative} \\
&= \det(A - tI_n)\det(Q^{-1}Q) \\
&= p_A(t).
\end{aligned}$$

□

Definition. Let V be a vector space over $\mathbb{F}$. A linear transformation $T : V \to V$ is called a **linear operator** on V.

$L(V)$ denotes the set of all linear operators on V.

Definition. Let V be a vector space over $\mathbb{F}$ and let $T \in L(V)$. An **eigenvector** of T is any nonzero vector $v \in V$ satisfying $T(v) \in \operatorname{span}(v)$. The (unique) scalar $\lambda \in \mathbb{F}$ satisfying $T(v) = \lambda v$ is the **eigenvalue** of T corresponding to v.

Definition. Suppose $T \in L(V)$. Given an eigenvalue λ of T, the set

$$\begin{aligned} E_\lambda &= \{v \in V : T(v) = \lambda v\} \\ &= \{\text{eigenvectors of } T \text{ corresponding to } \lambda\} \cup \{0\}. \end{aligned}$$

is called the **eigenspace** of T corresponding to λ.

Example. Consider $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ given by $D(f) = f'$. Every $\lambda \in \mathbb{R}$ is an eigenvalue of D, since $D(e^{\lambda x}) = \lambda e^{\lambda x}$.

For the rest of this term we focus on the case when V is finite-dimensional. In this case, can we define the characteristic polynomial of a linear operator $T \in L(V)$? We could pick an ordered basis α for V, define $A = [T]_\alpha$, and take $p_A(t)$. But what if we picked a different ordered basis β; would that give us a different characteristic polynomial?

Lemma 5.3. *Suppose V is finite-dimensional and $T \in L(V)$. Let α and β be two ordered bases for V and let $A = [T]_\alpha$ and $B = [T]_\beta$. Then $p_A(t) = p_B(t)$.*

Proof. Let $Q = [I_V]_\alpha^\beta$ be the change-of-coordinates matrix from α to β. By the last Theorem from Feb 14,

$$A = Q^{-1}BQ.$$

Thus A is similar to B, so $p_A(t) = p_B(t)$ by "Exercise 12." □

This lemma justifies the following definition.

Definition. Let $T \in L(V)$ where V is finite-dimensional. The **characteristic polynomial** of T is the characteristic polynomial of $[T]_\alpha$ for any ordered basis α for V. We denote the polynomial by $p_T(t)$.

If $T \in L(V)$ and V is a vector space over $\mathbb{F}$, then $p_T(t) \in \mathbb{F}[t]$. We are interested in the eigenvalues of T (the roots of $p_T(t)$ which belong to $\mathbb{F}$). We can also ask about how $p_T(t)$ factors in $\mathbb{F}[t]$. These issues are related: λ is a root iff $t - \lambda$ is a factor. Both issues depend sensitively on $\mathbb{F}$.

Example. Suppose $p_A(t) = t^4 + t^3 + t^2 + t + 1$.

- If $\mathbb{F} = \mathbb{Q}$, then $p_A(t)$ does not factor. (It is *irreducible.*) A has no eigenvalues.
- If $\mathbb{F} = \mathbb{R}$, then

$$p_A(t) = \left(t^2 + \frac{\sqrt{5}+1}{2}t + 1\right)\left(t^2 - \frac{\sqrt{5}-1}{2}t + 1\right)$$

  A again has no eigenvalues.

- If $\mathbb{F} = \mathbb{C}$, then
$$p_A(t) = (t - a_1)(t - a_2)(t - a_3)(t - a_4)$$
where $a_k = \operatorname{cis}(2k\pi/5) = \cos(2k\pi/5) + i\sin(2k\pi/5) \in \mathbb{C}$ for $k = 1, \ldots, 4$. A has 4 eigenvalues. In this case we say that $p_A(t)$ *splits*.
- If $\mathbb{F} = \mathbb{Z}_5$, then
$$p_A(t) = (t-1)^4 = (t-1)(t-1)(t-1)(t-1).$$
$p_A(t)$ again splits. A has one eigenvalue, of *multiplicity* 4.

Definition. A polynomial $f(t) \in \mathbb{F}[t]$ **splits over** $\mathbb{F}$ if there exist scalars $c, a_1, \ldots, a_n \in \mathbb{F}$ (not necessarily distinct) such that
$$f(t) = c(t - a_1)(t - a_2)\cdots(t - a_n).$$

Definition. Suppose V is finite-dimensional, $T \in L(V)$, and λ is an eigenvalue of T. The **multiplicity** of λ is the maximum value k such that $(t-\lambda)^k$ is a factor of $p_T(t)$.

Theorem 5.4. *Suppose V is finite-dimensional, $T \in L(V)$, λ is an eigenvalue of T, and m is the multiplicity of λ. Then $\dim(E_\lambda) \leq m$.*

Proof. Let $d = \dim(E_\lambda)$. Let $\alpha = (v_1, \ldots, v_d)$ be an ordered basis for E_λ. Extend α to an ordered basis $\beta = (v_1, \ldots, v_d, v_{d+1}, \ldots, v_n)$ for V. Let $A = [T]_\beta$, so $p_T(t) = p_A(t)$.

Observe that for $i = 1, \ldots, d$,
$$\begin{aligned} T(v_i) &= \lambda v_i \qquad\qquad \text{(because } v_i \in E_\lambda\text{)} \\ &= 0v_1 + \cdots + 0v_{i-1} + \lambda v_i + 0v_{i+1} + \cdots 0v_d + 0v_{d+1} + \cdots + 0v_n. \end{aligned}$$

Hence
$$A = \left(\begin{array}{cccc|ccc} \lambda & 0 & \cdots & 0 & * & \cdots & * \\ 0 & \lambda & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots & * & \cdots & * \\ 0 & 0 & \cdots & \lambda & * & \cdots & * \\ \hline 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & * & \cdots & * \end{array}\right) = \begin{pmatrix} \lambda I_d & B \\ O & C \end{pmatrix}$$

for some matrices of the appropriate dimensions. Thus
$$\begin{aligned} p_T(t) = \det(A - tI_n) &= \det\begin{pmatrix} (\lambda - t)I_d & B \\ O & C - tI_{n-d} \end{pmatrix} \\ &= \det((\lambda - t)I_d)\cdot\det(C - tI_{n-d}) \qquad \text{by A5Q6} \\ &= (\lambda - t)^d \det(I_d)\cdot\det(C - tI_{n-d}) \qquad \text{since } \det(cA) = c^n\det(A) \text{ if } A \text{ is } n\times n \\ &= (\lambda - t)^d \cdot p_C(t). \end{aligned}$$

This proves that $(t-\lambda)^d$ divides $p_T(t)$. Since m, the multiplicity of λ, is by definition the *largest* value such that $(t-\lambda)^m$ divides $p_T(t)$, we have proved that $d \leq m$, i.e., $\dim(E_\lambda) \leq m$. $\square$

MATH 146 March 30 Lecture notes

Example. Let $A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$, considered as a real matrix (i.e., over $\mathbb{R}$).

$$\begin{aligned} p_A(t) \;&=\; \det(A - tI_3) \;=\; \det \begin{pmatrix} 4-t & 0 & 1 \\ 2 & 3-t & 2 \\ 1 & 0 & 4-t \end{pmatrix} \;=\; (3-t)\det \begin{pmatrix} 4-t & 1 \\ 1 & 4-t \end{pmatrix} \\ &=\; (3-t)((4-t)^2 - 1) \;=\; (3-t)(t^2 - 8t + 15) \;=\; -(t-3)^2(t-5). \end{aligned}$$

Thus $p_A(t)$ splits over $\mathbb{R}$, and has two real eigenvalues $\lambda = 3, 5$ of multiplicities $m = 2, 1$ respectively. Let's find the corresponding eigenspaces, their dimensions, and a basis for each.

$\boxed{\lambda = 3}$

$$E_3 = N(A - 3I_3) = N \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix}$$

By inspection, $\operatorname{rank}(A - 3I_3) = 1$, so $\dim(E_3) = \operatorname{nullity}(A - 3I_3) = 2$. To get a basis for E_3, solve the system $(A - 3I_3)x = 0$. The augmented matrix for this system is

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{in RREF.}$$

Writing $x = (x_1, x_2, x_3)$ and introducing parameters s, t for the non-leading variables x_2, x_3, we solve the above system to find

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = s \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad s, t \in \mathbb{R}.$$

Thus a basis for the solution set is $\{v_1, v_2\}$ where $v_1 = (0, 1, 0)$, $v_2 = (-1, 0, 1)$.

$\boxed{\lambda = 5}$

$$E_5 = N(A - 5I_3) = N \begin{pmatrix} -1 & 0 & 1 \\ 2 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix}$$

By inspection, $\operatorname{rank}(A - 5I_3) = 2$, so $\dim(E_5) = \operatorname{nullity}(A - 5I_3) = 1$.

(Note that $\dim(E_5)$ automatically equals 1, since $1 \le \dim(E_5) \le$ (multiplicity of 5) $= 1$.)

To get a basis for E_5, solve the system $(A - 5I_3)x = 0$. The augmented matrix for this system is

$$\left(\begin{array}{ccc|c} -1 & 0 & 1 & 0 \\ 2 & -2 & 2 & 0 \\ 1 & 0 & -1 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \text{in RREF.}$$

By solving, we find that a basis for the solution set is $\{v_3\}$ where $v_3 = (1, 2, 1)$.

In this example, we can see that $v_3 \notin \operatorname{span}(v_1, v_2)$, since v_1 and v_2 both lie in the plane defined by $x + z = 0$ but v_2 does not. Hence v_1, v_2, v_3 are linearly independent and so $\beta = (v_1, v_2, v_3)$ is an ordered basis for $\mathbb{R}^3$. I will prove on Wednesday that, more generally, whenever we combine bases for different eigenspaces, the resulting set is linearly independent.

Example. Let's do the same thing for the matrix $B = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$.

$$p_B(t) = \det(B - tI_3) = \det \begin{pmatrix} 3-t & 1 & 0 \\ 0 & 3-t & 0 \\ 0 & 0 & 5-t \end{pmatrix} = (3-t)^2(5-t) = -(t-3)^2(t-5).$$

The same as $p_A(t)$; hence the same eigenvalues, with the same multiplicities. Let's find the eigenspaces and their dimensions.

$\boxed{\lambda = 3}$

$$E_3 = N(B - 3I_3) = N \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

By inspection, $\operatorname{rank}(B - 3I_3) = 2$, so $\dim(E_3) = \operatorname{nullity}(B - 3I_3) = 1$. To get a basis for E_3, solve the system $(B - 3I_3)x = 0$. A basis for the solution set is $\{v_1\}$ where $v_1 = (1, 0, 0)$.

$\boxed{\lambda = 5.}$

Since 5 has multiplicity 1, we know that $\dim(E_5) = 1$. By inspection, $v_2 = (0, 0, 1)$ is an eigenvector for 5, so $\{v_2\}$ is a basis for E_5.

Note that in the second example, the bases for the two eigenspaces, when merged, do NOT form a basis for $\mathbb{R}^3$. But their union is at least linearly independent. As mentioned before, this is always true. As preparation for this Theorem (which I will prove on Wednesday), we need the following Lemma.

Lemma 5.5. *Suppose V is finite-dimensional, $T \in L(V)$, and $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T. If $x_1, \ldots, x_k$ are eigenvectors corresponding to $\lambda_1, \ldots, \lambda_k$, then $\{x_1, \ldots, x_k\}$ is linearly independent.*

Proof. By induction on k. When $k = 1$ the claim is obvious. Suppose $k > 1$ and the claim is true for lower values. In particular, $\{x_1, \ldots, x_{k-1}\}$ is linearly independent. Now let's prove that $\{x_1, \ldots, x_k\}$ is linearly independent. Assume

$$(*) \qquad a_1x_1 + a_2x_2 + \cdots + a_{k-1}x_{k-1} + a_kx_k = 0.$$

Apply T to both sides to get

$$a_1T(x_1) + a_2T(x_2) + \cdots + a_{k-1}T(x_{k-1}) + a_kT(x_k) = 0$$

which simplifies to

$$(**) \qquad a_1\lambda_1x_1 + a_2\lambda_2x_2 + \cdots + a_{k-1}\lambda_{k-1}x_{k-1} + a_k\lambda_kx_k = 0.$$

Subtract λ_k times $(*)$ from $(**)$ to get

$$a_1(\lambda_1 - \lambda_k)x_1 + a_2(\lambda_2 - \lambda_k)x_2 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = 0.$$

Since $\{x_1, \ldots, x_{k-1}\}$ is linearly independent, we get

$$a_1(\lambda_1 - \lambda_k) = a_2(\lambda_2 - \lambda_k) = \cdots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Since $\lambda_i \neq \lambda_k$ for $i < k$, we get

$$a_1 = a_2 = \cdots = a_{k-1} = 0.$$

Thus $(*)$ implies $a_kx_k = 0$. Since $x_k \neq 0$ (it is an eigenvector), it follows that $a_k = 0$. So we've proved $a_1 = a_2 = \cdots = a_k = 0$, which proves $\{x_1, x_2, \ldots, x_k\}$ is linearly independent. $\square$

Recall from Monday:

Lemma 5.5. *Suppose V is finite-dimensional, $T \in L(V)$, and $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of T. If $x_1, \dots, x_k$ are eigenvectors corresponding to $\lambda_1, \dots, \lambda_k$, then $\{x_1, \dots, x_k\}$ is linearly independent.*

Corollary. *If $V, T, \lambda_1, \dots, \lambda_k$ are as in Lemma 5.5 and $x_i \in E_{\lambda_i}$ for $i = 1, \dots, k$, then*

$$x_1 + \cdots + x_k = 0 \implies x_1 = \cdots = x_k = 0.$$

Proof. The nonzero vectors (if any) among $x_1, \dots, x_k$ are eigenvectors for distinct eigenvalues, so are linearly independent by Lemma 5.5, yet they sum to 0. □

Theorem 5.6. *Suppose V is finite-dimensional, $T \in L(V)$, and $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of T. If for each $i = 1, \dots, k$, B_i is a basis for E_{λ_i}, then*

(1) *$B_i \cap B_j = \varnothing$ for $i \neq j$, and*
(2) *$B_1 \cup B_2 \cup \cdots \cup B_k$ is linearly independent.*

Proof. (1) is easy: a nonzero vector cannot be an eigenvector for two different eigenvalues.

(2) For each i let $d_i = \dim(E_{\lambda_i})$ and write $B_i = \{v_1^i, v_2^i, \dots, v_{d_i}^i\}$. Then let

$$B = B_1 \cup \cdots \cup B_k = \{\underbrace{v_1^1, v_2^1, \dots, v_{d_1}^1}_{B_1}, \underbrace{v_1^2, v_2^2, \dots, v_{d_2}^2}_{B_2}, \dots, \underbrace{v_1^k, v_2^k, \dots, v_{d_k}^k}_{B_k}\}.$$

Assume

$$(\dagger) \qquad \underbrace{a_1^1 v_1^1 + a_2^1 v_2^1 + \cdots + a_{d_1}^1 v_{d_1}^1}_{x_1} + \underbrace{a_1^2 v_1^2 + a_2^2 v_2^2 + \cdots + a_{d_2}^2 v_{d_2}^2}_{x_2} + \cdots + \underbrace{a_1^k v_1^k + a_2^k v_2^k + \cdots + a_{d_k}^k v_{d_k}^k}_{x_k} = 0.$$

For $i = 1, \dots, k$ let $x_i = \sum_{j=1}^{d_i} a_j^i v_j^i$ and note that $x_i \in E_{\lambda_i}$. Thus our assumption can be summarized as

$$x_1 + x_2 + \cdots + x_k = 0.$$

The Corollary then gives $x_1 = x_2 = \cdots = x_k = 0$. Using linear independence of each B_i, we get that all the coefficients in ($\dagger$) equal 0. □

Definition.

(1) A square matrix is **diagonal** if every entry <u>not</u> on the diagonal is 0.
(2) Suppose V is finite-dimensional and $T \in L(V)$. T is **diagonalizable** if there exists an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix.
(3) Suppose $A \in \mathsf{M}_{n\times n}(\mathbb{F})$. A is **diagonalizable** (over $\mathbb{F}$) if A is similar (over $\mathbb{F}$) to a diagonal matrix D. I.e., there exists an invertible $Q \in \mathsf{M}_{n\times n}(\mathbb{F})$ such that $Q^{-1}AQ = D$.

Exercise. For any $A \in M_{n\times n}(\mathbb{F})$, A is diagonalizable (over $\mathbb{F}$) iff L_A is diagonalizable.

How can we tell if an operator (or a matrix) is diagonalizable? The rest of today's lecture is devoted to this question. Suppose $T \in L(V)$ is diagonalizable. Let $\beta = (v_1, \dots, v_n)$ be an ordered basis for V such that $[T]_\beta$ is a diagonal matrix, i.e.,

$$[T]_\beta = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

This implies that $T(v_1) = a_{11}v_1$, $T(v_2) = a_{22}v_2$,..., $T(v_n) = a_{nn}v_n$. In otherwords, each v_i is an eigenvector of T (and a_{ii} is its corresponding eigenvalue). Conversely, if T has an ordered basis $\beta = (v_1, \dots, v_n)$ consisting of eigenvectors, then it is easy to see that $[T]_\beta$ is a diagonal matrix. This proves

Theorem 5.7. *Suppose V is a finite-dimensional vector space over $\mathbb{F}$ and $T \in L(V)$. T is diagonalizable iff V has an ordered basis consisting of eigenvectors of T.*

Here is a more useful theorem:

Theorem 5.8. *Suppose V is finite-dimensional, say $\dim(V) = n$, and $T \in L(V)$. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T and let $m_1, \dots, m_k$ be their multiplicities. T is diagonalizable iff*

(a) *$p_T(t)$ splits, i.e., $p_T(t) = (-1)^n(t-\lambda_1)^{m_1} \cdots (t-\lambda_k)^{m_k}$, and*

(b) *For each $i = 1, \dots, k$, $\dim(E_{\lambda_i}) = m_i$.*

Proof. ($\Leftarrow$) Assume that (a) and (b) hold. For each $i = 1, \dots, k$ let B_i be a basis for E_{λ_i}; thus $|B_i| = m_i$ by (b). Let $B = B_1 \cup \cdots \cup B_k$. B is linearly independent by Theorem 5.6(2), and $|B| = m_1 + \cdots + m_k = n$ by Theorem 5.6(1). Hence B is a basis for V consisting of eigenvectors for T. Hence T is diagonalizable by Theorem 5.7.

($\Rightarrow$) Assume T is diagonalizable. By Theorem 5.7, T has a basis $B = \{v_1, \dots, v_n\}$ consisting of eigenvectors of T. Each v_i belongs to $E_{\lambda_1} \cup \cdots \cup E_{\lambda_k}$. For each $i = 1, \dots, k$, let $B_i = B \cap E_{\lambda_i}$ and $n_i = |B_i|$. Then

$$(*) \qquad n_1 + \cdots + n_k = |B| = n.$$

For each i let m_i be the multiplicity of λ_i. Hence

$$(t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2} \cdots (t-\lambda_k)^{m_k}$$

is a factor of $p_T(t)$, which has degree n. So

$$(**) \qquad m_1 + \cdots + m_k \le n.$$

Finally, for each $i = 1, \dots, k$ we have

$$(\dagger) \qquad n_i \le \dim(E_{\lambda_i}) \le m_i,$$

where the first $\le$ is due to the fact that B_i is linearly independent in E_{λ_i}, while the second $\le$ follows from Theorem 5.4. Combining $(*)$, $(**)$ and $(\dagger)$ we get $n_i = \dim(E_{\lambda_i}) = m_i$ for all i, proving (b), and since $m_1 + \cdots + m_k = n$ we must also have (a). □

(The proof also shows that each B_i is a basis for E_{λ_i}.)

Recall from Wednesday's lecture:

Theorem 5.8. *Suppose V is finite-dimensional, say $\dim(V) = n$, and $T \in L(V)$. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T and let $m_1, \ldots, m_k$ be their multiplicities. T is diagonalizable iff*

(a) *$p_T(t)$ splits, i.e., $p_T(t) = (-1)^n (t-\lambda_1)^{m_1} \cdots (t-\lambda_k)^{m_k}$, and*
(b) *For each $i = 1, \ldots, k$, $\dim(E_{\lambda_i}) = m_i$.*

Example. Let $A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$. On March 30 we saw that $p_A(t) = -(t-3)^2(t-5)$ and $\dim(E_3) = 2$ and $\dim(E_5) = 1$. Hence A is diagonalizable by Theorem 5.8.

A basis for E_3 was $\{v_1, v_2\}$ where $v_1 = (0,1,0)$ and $v_2 = (-1,0,1)$. A basis for E_5 was $\{v_3\}$ where $v_3 = (1,2,1)$. Hence by the proof of Theorem 5.7, $\beta = (v_1, v_2, v_3)$ is an ordered basis for $\mathbb{R}^3$ and

$$[L_A]_\beta = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} =: D.$$

To find an invertible matrix Q such that $Q^{-1}AQ = D$, let σ be the standard ordered basis for $\mathbb{R}^3$. Note that $D = [L_A]_\beta = [I]_\sigma^\beta \cdot [L_A]_\sigma \cdot [I]_\beta^\sigma = ([I]_\beta^\sigma)^{-1} \cdot A \cdot [I]_\beta^\sigma$, so we can take $Q = [I]_\beta^\sigma$, which is simply the matrix whose columns are v_1, v_2, v_3 expressed in standard form. That is, if

$$Q = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$$

then

$$Q^{-1}AQ = D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Example. Let $B = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \in M_{3\times 3}(\mathbb{R})$. On March 30 we saw that B has the same characteristic polynomial, eigenvalues, and multiplicities as A. But this time $\dim(E_3) = 1$. So B is not diagonalizable by Theorem 5.8.

Here is a useful observation.

Corollary 5.9. *Suppose $T \in L(V)$ with $\dim(V) = n$. If T has n <u>distinct</u> eigenvalues, then T is diagonalizable.*

Proof. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues. Then $p_T(t) = (-1)^n (t-\lambda_1)\cdots(t-\lambda_n)$, so $p_T(t)$ splits and each multiplicity equals 1. Then $1 \le \dim(E_{\lambda_i}) \le 1$ by Theorem 5.4; so we have equality for all i. $\square$

Example. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$. Is A diagonalizable? What if we consider $A \in M_{2\times 2}(\mathbb{C})$?

If a matrix A is diagonalizable, it is easy to explicitly compute A^k for any k. Here is how.

(1) If D is a diagonal matrix, say $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$, then

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

Indeed, for each i, $De_i = \lambda_i e_i$, so $D^2 e_i = D(\lambda_i e_i) = \lambda_i (De_i) = (\lambda_i)^2 e_i$, etc.

(2) If A is diagonalizable, then there exists an invertible Q such that $Q^{-1}AQ = D$ where D is diagonal. We can rewrite this as $A = QDQ^{-1}$. So

$$\begin{aligned} A^k = (QDQ^{-1})^k &= (QDQ^{-1})(QDQ^{-1})\cdots(QDQ^{-1}) \\ &= QD(Q^{-1}Q)D(Q^{-1}Q)D\cdots(Q^{-1}Q)DQ^{-1} \\ &= QD^kQ^{-1}. \end{aligned}$$

Example. Let $A = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$. We've seen that A is diagonalizable, and if $Q = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$ then

$$Q^{-1}AQ = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} =: D.$$

Thus $A = QDQ^{-1}$ and hence $A^k = QD^kQ^{-1}$ for any k. To compute A^k explicitly, we need to know Q^{-1}:

$$\begin{aligned} (Q|I_3) &= \left(\begin{array}{ccc|ccc} 0 & -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{array}\right) \\ &\to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 2 & 1 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right) = (I_3|Q^{-1}). \end{aligned}$$

Hence $Q^{-1} = \begin{pmatrix} -1 & 1 & -1 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$, so

$$\begin{aligned} A^k = QD^kQ^{-1} &= \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3^k & 0 & 0 \\ 0 & 3^k & 0 \\ 0 & 0 & 5^k \end{pmatrix} \begin{pmatrix} -1 & 1 & -1 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & -3^k & 5^k \\ 3^k & 0 & 2\cdot 5^k \\ 0 & 3^k & 5^k \end{pmatrix} \begin{pmatrix} -1 & 1 & -1 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(3^k+5^k) & 0 & \frac{1}{2}(5^k-3^k) \\ 5^k-3^k & 3^k & 5^k-3^k \\ \frac{1}{2}(5^k-3^k) & 0 & \frac{1}{2}(3^k+5^k) \end{pmatrix}. \end{aligned}$$