MA 35300–001–60150
MA 35300–002–60151 **GOINS**

FINAL EXAM STUDY GUIDE

The following sections will be covered on the Final Exam:

- §1.1: Introduction
- §1.2: Vector Spaces
- §1.3: Subspaces
- §1.4: Linear Combinations and Systems of Linear Equations
- §1.5: Linear Dependence and Linear Independence
- §1.6: Bases and Dimension
- §1.7: Maximal Linearly Independent Subsets

- §2.1: Linear Transformations, Null Spaces, and Ranges
- §2.2: The Matrix Representation of a Linear Transformation
- §2.3: Composition of Linear Transformations and Matrix Multiplication
- §2.4: Invertibility and Isomorphisms
- §2.5: The Change of Coordinate Matrix

- §3.1: Elementary Matrix Operations and Elementary Matrices
- §3.2: The Rank of a Matrix and Matrix Inverses
- §3.3: Systems of Linear Equations – Theoretical Aspects
- §3.4: Systems of Linear Equations – Computational Aspects

- §4.1: Determinants of Order 2
- §4.2: Determinants of Order n
- §4.3: Properties of Determinants
- §4.4: Summary – Important Facts about Determinants

- §5.1: Eigenvalues and Eigenvectors
- §5.2: Diagonalizability
- §5.3: Matrix Limits

- §6.1: Inner Products and Norms
- §6.2: The Gram-Schmidt Orthogonalization Process and Orthogonal Complements
- §6.3: The Adjoint of a Linear Operator
- §6.4: Normal and Self-Adjoint Operators

§1.1: Introduction.

> **Parallelogram Law for Vector Addition.** The sum of two vectors x and y that act at the same point P is the vector $x+y$ beginning at P that is represented by the diagonal of a parallelogram having x and y as adjacent sides.

§1.2: Vector Spaces. We say that a triple $(\mathsf{V},\,+,\,\cdot)$ is a *vector space over* F if we have a nonempty set V, an operation $+ : \mathsf{V} \times \mathsf{V} \to \mathsf{V}$ called *vector addition*, and an operation $\cdot : F \times \mathsf{V} \to \mathsf{V}$ called *scalar multiplication* satisfying the following eight axioms:

(VS 1) *Commutativity of Vector Addition:* For all x and y in V, we have $x + y = y + x$.
(VS 2) *Associativity of Vector Addition:* For all x, y, and z in V, we have $(x+y)+z = x+(y+z)$.
(VS 3) *Vector Additive Identity:* There exists O in V such that $x + O = x$ for each x in V. We call $O = 0\,x$ the *zero element*.
(VS 4) *Vector Additive Inverse:* For each x in V, there is a y in V such that $x + y = O$. Many times, we denote this element by $-x = (-1)\,x$.
(VS 5) *Scalar Multiplicative Identity:* For each vector x we have $x = 1\,x$.
(VS 6) *Associativity of Scalar Multiplication:* For all a and b in F as well as x in V, we have $(a\,b)\,x = a\,(b\,x)$.
(VS 7) *Distributivity for Vector Addition:* For all a in F as well as x and y in V, we have $a\,(x+y) = a\,x + a\,y$.
(VS 8) *Distributivity for Scalar Multiplication:* For all a in F and b as well as x in V, we have $(a+b)\,x = a\,x + b\,x$.

§1.3: Subspaces.

> **Theorem 1.3.** Assume that $(\mathsf{V},\,+,\,\cdot)$ is a vector space over a field F. The following are equivalent for a subset $\mathsf{W} \subseteq \mathsf{V}$.
>
> i. $(\mathsf{W},\,+,\,\cdot)$ a vector space over F.
> ii. The following three conditions hold for the operations "vector addition" $+$ and "scalar multiplication" $\cdot$ defined on V:
>   (a) *Zero Element:* $O \in \mathsf{W}$.
>   (b) *Closure under Vector Addition:* $x + y \in \mathsf{W}$ whenever $x \in \mathsf{W}$ and $y \in \mathsf{W}$.
>   (c) *Closure under Scalar Multiplication:* $c\,x \in \mathsf{W}$ whenever $c \in F$ and $x \in \mathsf{W}$.

Any subset $\mathsf{W} \subseteq \mathsf{V}$ satisfying either of the equivalent statements above is called a *subspace of* V.

§1.4: Linear Combinations and Systems of Linear Equations. Let $(\mathsf{V},\,+,\,\cdot)$ is a vector space over a field F, and choose a finite subset $S = \{u_1,\,u_2,\,\ldots,\,u_n\}$ of V. For any scalars $a_1,\,a_2,\,\ldots,\,a_n \in F$, we call the vector

$$v = a_1\,u_1 + a_2\,u_2 + \cdots + a_n\,u_n \in \mathsf{V}$$

a *linear combination of vectors from* S and the a_i are called the *coefficients*.

> **Theorem 1.5.** Let $(\mathsf{V}, +, \cdot)$ be a vector space over a field F, and choose a finite subset $S = \{u_1, u_2, \ldots, u_n\}$ of V. The collection of linear combinations of vectors from S, namely the subset
> $$\mathsf{W} = \left\{ v \in \mathsf{V} \;\middle|\; v = a_1\, u_1 + a_2\, u_2 + \cdots + a_n\, u_n \text{ for some } a_1, a_2, \ldots, a_n \in F \right\}$$
> is a subspace of V.

We often define $\mathsf{W} = \operatorname{span}(S)$ as the *span of* S. When $S = \emptyset$ has no elements, we define $\operatorname{span}(S) = \{O\}$ as the zero subspace. We say that S *generates* W.

§1.5: Linear Dependence and Linear Independence. Let $(\mathsf{V}, +, \cdot)$ be a vector space over a field F, and choose a finite subset $S = \{u_1, u_2, \ldots, u_n\}$ of V. We say that S *is a linearly dependent set* if there exist scalars $a_1, a_2, \ldots, a_n \in F$ not all zero such that $a_1\, u_1 + a_2\, u_2 + \cdots + a_n\, u_n = O$. Such an equation is called a *nontrivial expression of the zero vector* O. Otherwise, we say that S *is a linearly independent set.*

> **Theorem 1.7.** Let $T = \{u_1, u_2, \ldots, u_n\}$ be a subset of a vector space V which is linearly indendent. Choose a vector $v \in \mathsf{V}$ such that $v \notin T$, and denote $\mathsf{W} = \operatorname{span}(S)$ as the span of the set $S = T \cup \{v\}$. Then $v \in \operatorname{span}(T)$ if and only if S is a linearly dependent set.

§1.6: Bases and Dimension.

> **Theorem 1.8.** Let $(\mathsf{V}, +, \cdot)$ be a vector space over a field F. Given a finite subset $\beta = \{u_1, u_2, \ldots, u_n\} \subseteq \mathsf{V}$, the following are equivalent:
>
> i. We have the two properties
>   a. $\mathsf{V} = \operatorname{span}(\beta)$, that is, β spans V; and
>   b. β is a linearly independent set.
>
> ii. Each $v \in \mathsf{V}$ can be uniquely expressed as a linear combination of vectors from β, that is, $v = a_1\, u_1 + a_2\, u_2 + \cdots + a_n\, u_n$ for some unique scalars $a_1, a_2, \ldots, a_n \in F$.

If either of these properties are true, we say that β a *basis for* V.

> **Theorem 1.9.** Let $(\mathsf{V}, +, \cdot)$ be a vector space over a field F. Assume that $\mathsf{V} = \operatorname{span}(S)$ for some finite set S.
>
> - V has a finite basis β contained in S.
> - Every basis β of V has the same number of elements.

For any basis β of V, we call $\dim(\mathsf{V}) = |\beta|$ the *dimension* of V.

- Let $\mathsf{V} = \{O\}$ denote the zero space. Then $\beta = \emptyset$ is a basis for V, so that $\dim(\mathsf{V}) = 0$.

- Let $\mathsf{V} = \mathbb{R}^2$ denote the real plane. For any $a, b \in \mathbb{R}$ not both zero, and denote the two vectors
$$u_1 = \frac{1}{\sqrt{a^2 + b^2}} \begin{pmatrix} a \\ b \end{pmatrix} \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{a^2 + b^2}} \begin{pmatrix} -b \\ a \end{pmatrix}.$$

Then $\beta = \{u_1,\, u_2\}$ is a basis for V, so that $\dim(\mathsf{V}) = 2$.

- Let $\mathsf{V} = \mathbb{R}^3$ denote three space. For any $a,\, b,\, c \in \mathbb{R}$ such that a and b are not both zero, and denote the three vectors

$$u_1 = \frac{1}{\sqrt{a^2+b^2+c^2}} \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

$$u_2 = \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} -b \\ a \\ 0 \end{pmatrix},$$

$$u_3 = \frac{1}{\sqrt{a^2+b^2}\sqrt{a^2+b^2+c^2}} \begin{pmatrix} -a\,c \\ -b\,c \\ a^2+b^2 \end{pmatrix}.$$

  Then $\beta = \{u_1,\, u_2,\, u_3\}$ is a basis for V, so that $\dim(\mathsf{V}) = 3$.

- Consider the collection $\mathsf{V} = \mathsf{F}^m$ of m-dimensional vectors over a field F. Then $\beta = \{e_1,\, e_2,\, e_3,\, \ldots,\, e_m\}$ is a basis for $\mathsf{V} = \mathsf{F}^m$, where e_i is that m-dimensional vector whose only nonzero entry is a 1 on the ith row. In particular, $\dim(\mathsf{F}^m) = m$.

- Consider the collection $\mathsf{V} = \mathsf{M}_{m\times n}(F)$ of $m \times n$ matrices over a field F. Then $\beta = \{E^{ij} \mid 1 \le i \le m,\ 1 \le j \le n\}$ is a basis for $\mathsf{V} = \mathsf{M}_{m\times n}(F)$, where E^{ij} is that $m \times n$ matrix whose only nonzero entry is a 1 in the ith row and jth column. In particular, $\dim\big(\mathsf{M}_{m\times n}(F)\big) = m\,n$.

§1.7: Maximal Linearly Independent Subsets.

Theorems 1.11, 1.12. Let $(\mathsf{V},\, +,\, \cdot)$ be a vector space over a field F, and let $\mathsf{W} \subseteq \mathsf{V}$ be a subspace. Assume that $\mathsf{V} = \operatorname{span}(\beta)$ for some basis $\beta = \{u_1,\, u_2,\, \ldots,\, u_n\}$.

- $\mathsf{W} = \operatorname{span}(T)$ has a finite basis T.
- $\dim(\mathsf{W}) \le \dim(\mathsf{V})$.
- $\dim(\mathsf{W}) = \dim(\mathsf{V})$ if and only if $\mathsf{W} = \mathsf{V}$.

§2.1: Linear Transformations, Null Spaces, and Ranges. Let $(\mathsf{V},\, \oplus,\, \odot)$ and $(\mathsf{W},\, \boxplus,\, \boxdot)$ be two vector spaces over a common field F. We say that a map $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is *well-defined* if $\mathsf{T}(x_1) = \mathsf{T}(x_2)$ in W whenever $x_1 = x_2$ in V. In this case, we say that V is the *domain* and W is the *codomain*.

Proposition. Given a well-defined map $\mathsf{T} : \mathsf{V} \to \mathsf{W}$, the following are equivalent.

i. For all $x, y \in \mathsf{V}$ and $c \in F$ we have
   (a) *Preserves Vector Addition:* $\mathsf{T}\big(x \oplus y\big) = \mathsf{T}(x) \boxplus \mathsf{T}(y)$, and
   (b) *Preserves Scalar Multiplication:* $\mathsf{T}\big(c \odot x\big) = c \boxdot \mathsf{T}(x)$.

ii. For any collection $v_1, v_2, \dots, v_n \in \mathsf{V}$ and $a_1, a_2, \dots, a_n \in F$ we have

$$T\big((a_1 \odot v_1) \oplus (a_2 \odot v_2) \oplus \cdots \oplus (a_n \odot v_n)\big)$$
$$= \big(a_1 \boxdot \mathsf{T}(v_1)\big) \boxplus \big(a_2 \boxdot \mathsf{T}(v_2)\big) \boxplus \cdots \boxplus \big(a_n \boxdot \mathsf{T}(v_n)\big).$$

A well-defined map $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ satisfying either of these equivalent conditions is said to be a *linear transformation.*

Theorem 2.1, 2.4. Let $\big(\mathsf{V}, \oplus, \odot\big)$ and $\big(\mathsf{W}, \boxplus, \boxdot\big)$ be two vector spaces over a common field F, and say that $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is a linear transformation.

- Let $O_{\mathsf{V}} \in \mathsf{V}$ and $O_{\mathsf{W}} \in \mathsf{W}$ denote the additive identity elements. Then $\mathsf{T}\big(O_{\mathsf{V}}\big) = O_{\mathsf{W}}$.
- The subset $\mathsf{N}(\mathsf{T}) = \left\{ x \in \mathsf{V} \;\middle|\; \mathsf{T}(x) = O_{\mathsf{W}} \right\}$ is a subspace of V. Moreover, the map $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is one-to-one if and only if $\mathsf{N}(\mathsf{T}) = \big\{O_{\mathsf{V}}\big\}$ is the zero subspace.
- The subset $\mathsf{R}(\mathsf{T}) = \left\{ u \in \mathsf{W} \;\middle|\; u = \mathsf{T}(x) \text{ for some } x \in \mathsf{V} \right\}$ is a subspace of W. Moreover, the map $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is onto if and only if $\mathsf{R}(\mathsf{T}) = \mathsf{W}$ is the entire space.

The subset $\mathsf{N}(\mathsf{T}) \subseteq \mathsf{V}$ is called the *null space* or the *kernel* of T. The subset $\mathsf{R}(\mathsf{T}) \subseteq \mathsf{W}$ is called the *range* or the *image* of T. If $\mathsf{V} = \operatorname{span}(\beta)$ and $\mathsf{W} = \operatorname{span}(\gamma)$ have finite bases $\beta \subseteq \mathsf{V}$ and $\gamma \subseteq \mathsf{W}$, then Theorems 1.11 and 1.12 say that both $\mathsf{N}(\mathsf{T}) \subseteq \mathsf{V}$ and $\mathsf{R}(\mathsf{T}) \subseteq \mathsf{W}$ also have finite bases. We define the *nullity of T* as the dimension of the null space, and the *rank of T* as the dimension of the range. That is,

$$\operatorname{nullity}(\mathsf{T}) = \dim \mathsf{N}(\mathsf{T}) \qquad \text{and} \qquad \operatorname{rank}(\mathsf{T}) = \dim \mathsf{R}(\mathsf{T}).$$

Theorem 2.3 (Dimension Theorem). Let $\big(\mathsf{V}, \oplus, \odot\big)$ and $\big(\mathsf{W}, \boxplus, \boxdot\big)$ be two finite-dimensional vector spaces over a common field F. For any linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{W}$, we have the equality

$$\operatorname{nullity}(\mathsf{T}) + \operatorname{rank}(\mathsf{T}) = \dim(\mathsf{V}).$$

Some texts call this result the *Rank-Nullity Theorem.*

§2.2: The Matrix Representation of a Linear Transformation. Let $\beta = \{v_1, v_2, \dots, v_n\}$ be a basis of a vector space $\big(\mathsf{V}, \oplus, \odot\big)$ over a field F. We will be concerned with the order of the elements v_i in β; then we call β an *ordered basis.* For each $x \in \mathsf{V}$, we can write

$$x = (a_1 \odot v_1) \oplus (a_2 \odot v_2) \oplus \cdots \oplus (a_n \odot v_n)$$

for some $a_1, a_2, \ldots, a_n \in F$. Theorem 1.8 asserts that this expression is *unique*. The *coordinate vector* of x relative to β is the well-defined n-dimensional vector

$$\big[x\big]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathsf{F}^n.$$

Theorem 2.14. Let $(\mathsf{V}, \oplus, \odot)$ and $(\mathsf{W}, \boxplus, \boxdot)$ be two vector spaces over a common field F having bases $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_m\}$, respectively. Say that $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is a linear transformation.

- Consider the map $\phi_\beta : \mathsf{V} \to \mathsf{F}^n$ which sends $x \mapsto \big[x\big]_\beta$. Then ϕ_β is a linear transformation.
- There is an $m \times n$ matrix $A = \big(A_{ij}\big) \in \mathsf{M}_{m \times n}(F)$ which makes the following diagram commute:

$$\begin{tikzcd} \mathsf{V} \arrow[r, "\mathsf{T}"] \arrow[d, "\phi_\beta"'] & \mathsf{W} \arrow[d, "\phi_\gamma"] \\ \mathsf{F}^n \arrow[r, "A"] & \mathsf{F}^m \end{tikzcd}$$

The map ϕ_β is called the *standard representation of V with respect to β*. The matrix $A = \big(A_{ij}\big)$ is denoted by $\big[\mathsf{T}\big]_\beta^\gamma$; it is the *matrix representation of T in the ordered bases β and γ*. Later, we will see that the nullity and rank of a linear transformation T are the same as the nullity and rank of the $m \times n$ matrix $\big[\mathsf{T}\big]_\beta^\gamma$.

§2.3: Composition of Linear Transformations and Matrix Multiplication. Say that we have two well-defined maps $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ and $\mathsf{U} : \mathsf{W} \to \mathsf{Z}$. Recall that the *composition* $\mathsf{U} \circ \mathsf{T} : \mathsf{V} \to \mathsf{Z}$ is that map which sends $x \in \mathsf{V}$ to $\big(\mathsf{U} \circ \mathsf{T}\big)(x) = \mathsf{U}\big[\mathsf{T}(x)\big] \in \mathsf{Z}$.

Theorem 2.7, 2.9, 2.11. Say that $(\mathsf{V}, \oplus, \odot)$, $(\mathsf{W}, \boxplus, \boxdot)$, and $(\mathsf{Z}, +, \cdot)$ are three vector spaces over a common field F.

- Let $\mathcal{L}(\mathsf{V}, \mathsf{W})$ denote the collection of all linear transformations $\mathsf{T} : \mathsf{V} \to \mathsf{W}$. Then $\mathcal{L}(\mathsf{V}, \mathsf{W})$ is a vector space over F.
- Say that $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ and $\mathsf{U} : \mathsf{W} \to \mathsf{Z}$ are linear transformations. Then $\mathsf{U} \circ \mathsf{T} : \mathsf{V} \to \mathsf{Z}$ is also a linear transformation.
- Say that V, W, and Z have finite bases $\beta = \{v_1, \ldots, v_n\}$, $\gamma = \{w_1, \ldots, w_m\}$, and $\alpha = \{z_1, \ldots, z_p\}$. Then we have the matrix product
$$\big[\mathsf{U} \circ \mathsf{T}\big]_\beta^\alpha = \big[\mathsf{U}\big]_\gamma^\alpha \, \big[\mathsf{T}\big]_\beta^\gamma.$$

§2.4: Invertibility and Isomorphisms.

Theorem 2.19. Let $(\mathsf{V}, \oplus, \odot)$ and $(\mathsf{W}, \boxplus, \boxdot)$ be two vector spaces over a common field F. The following are equivalent for a linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{W}$.

i. There exists a unique function $\mathsf{U} : \mathsf{W} \to \mathsf{V}$ such that the composition $\mathsf{T} \circ \mathsf{U} = \mathsf{I}_\mathsf{W} : \mathsf{W} \to \mathsf{W}$ is the identity map on W (that is, $\mathsf{I}_\mathsf{W}(y) = y$ for all $y \in \mathsf{W}$) and the composition $\mathsf{U} \circ \mathsf{T} = \mathsf{I}_\mathsf{V} : \mathsf{V} \to \mathsf{V}$ is the identity map on V (that is, $\mathsf{I}_\mathsf{V}(x) = x$ for all $x \in \mathsf{V}$).
ii. T is both one-to-one (that is, $\mathsf{T}(x_1) = \mathsf{T}(x_2)$ in W only when $x_1 = x_2$ in V) and onto (that is, given $y \in \mathsf{W}$, there exists $x \in \mathsf{V}$ such that $y = \mathsf{T}(x)$).
iii. $\operatorname{rank}(\mathsf{T}) = \dim(\mathsf{V}) = \dim(\mathsf{W})$ (assuming that V and W are finite-dimensional).

Any linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ satisfying either of these equivalent properties is said to be *invertible* or an *isomorphism*. The function $\mathsf{U} : \mathsf{W} \to \mathsf{V}$ is called the *inverse of T*, and we denote $\mathsf{T}^{-1} = \mathsf{U}$. In this case, we say that V and W are *isomorphic*, and we use the notation $\mathsf{V} \simeq \mathsf{W}$. The function $\mathsf{U} : \mathsf{W} \to \mathsf{V}$ is called the *inverse of T*, and we denote $\mathsf{T}^{-1} = \mathsf{U}$.

Theorem 2.17.

- If $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ is an isomorphism, so is its inverse $\mathsf{T}^{-1} : \mathsf{W} \to \mathsf{V}$.
- The binary relation $\simeq$ is an equivalence relation. That is,
  - *Reflexivity:* $\mathsf{V} \simeq \mathsf{V}$.
  - *Symmetry:* $\mathsf{V} \simeq \mathsf{W}$ if and only if $\mathsf{W} \simeq \mathsf{V}$.
  - *Transitivity:* If $\mathsf{V} \simeq \mathsf{W}$ and $\mathsf{W} \simeq \mathsf{Z}$, then $\mathsf{V} \simeq \mathsf{Z}$.

§2.5: The Change of Coordinate Matrix.

Theorem 2.18, 2.20. Let $(\mathsf{V}, \oplus, \odot)$ and $(\mathsf{W}, \boxplus, \boxdot)$ be two finite dimensional vector spaces over a common field F, and denote their bases as $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_m\}$, respectively.

- We have isomorphisms $\mathsf{V} \simeq \mathsf{F}^n$, $\mathsf{W} \simeq \mathsf{F}^m$, and $\mathcal{L}(\mathsf{V}, \mathsf{W}) \simeq \mathsf{M}_{m \times n}(F)$.
- Assume that there exists an isomorphism $\mathsf{T} : \mathsf{V} \to \mathsf{W}$. Then $\left[\mathsf{T}^{-1}\right]_\gamma^\beta$ is the inverse of the matrix $\left[\mathsf{T}\right]_\beta^\gamma$.

Theorem 2.22, 2.23. Let $(\mathsf{V}, +, \cdot)$ be a finite-dimensional vector space over field F, and say that $\beta = \{x_1, x_2, \ldots, x_n\}$ and $\beta' = \{x_1', x_2', \ldots, x_n'\}$ are two different bases for V. Denote the $n \times n$ matrix $Q = \left[\mathsf{I_V}\right]_{\beta'}^{\beta}$.

- The jth column of Q is the coordinate vector $\left[x_j'\right]_\beta$.
- Q is invertible, and its inverse is the $n \times n$ matrix $Q^{-1} = \left[\mathsf{I_V}\right]_{\beta}^{\beta'}$.
- $\left[v\right]_\beta = Q\left[v\right]_{\beta'}$ for any $v \in \mathsf{V}$.
- For any linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{V}$, there exist $n \times n$ matrices A and $B = Q^{-1} A\, Q$ which make the following diagram make sense:

$$\begin{tikzcd} \mathsf{F}^n \arrow[r, "B"] & \mathsf{F}^n \\ \mathsf{V} \arrow[u, "\phi_{\beta'}"] \arrow[r, "\mathsf{T}"] \arrow[d, "\phi_\beta"'] & \mathsf{V} \arrow[u, "\phi_{\beta'}"'] \arrow[d, "\phi_\beta"] \\ \mathsf{F}^n \arrow[r, "A"] & \mathsf{F}^n \end{tikzcd}$$

$$A = \left[\mathsf{T}\right]_\beta = \left[\mathsf{T}\right]_\beta^\beta$$

$$B = \left[\mathsf{T}\right]_{\beta'} = \left[\mathsf{T}\right]_{\beta'}^{\beta'} = Q^{-1} A\, Q$$

Remember that $\phi_\beta : \mathsf{V} \to \mathsf{F}^n$ is defined by $\phi_\beta(x) = \left[x\right]_\beta$; it is the *standard representation of* V *with respect to* β. The invertible $n \times n$ matrix Q is called the *change of coordinates matrix from from the* β'*-coordinates to the* β*-coordinates.* Note that if $\mathsf{V} = \mathsf{F}^n$ and $\beta = \{e_1, e_2, \ldots, e_n\}$ is the standard basis, then we can express the vectors for any other basis $\beta' = \{e_1', e_2', \ldots, e_n'\}$ in the form

$$e_j' = \left[e_j'\right]_\beta = \begin{pmatrix} Q_{1j} \\ Q_{2j} \\ \vdots \\ Q_{nj} \end{pmatrix} \qquad \text{for } 1 \leq j \leq n \qquad \Longrightarrow \qquad Q = \begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nn} \end{pmatrix}.$$

That is, the columns of Q are the vectors in $\beta' = \{e_1', e_2', \ldots, e_3'\}$. If A and B are $n \times n$ matrices, we say that B is *similar* to A if there exists an invertible $n \times n$ matrix Q such that such that $B = Q^{-1} A\, Q$. In this case, we write $B \simeq A$; we state without proof that this is an equivalence relation.

§3.1: Elementary Matrix Operations and Elementary Matrices. Let $A \in \mathsf{M}_{m \times n}(F)$ be an $m \times n$ matrix, that is, an array with m rows and n columns:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \qquad \text{where} \qquad A_{ij} \in F.$$

There are three types of operations we can perform on the rows (columns, respectively) of A:

- *Type 1:* Interchanging any two rows (columns, respectively) of A.
- *Type 2:* Multiplying any row (column, respectively) of A by a nonzero scalar.
- *Type 3:* Adding a scalar multiple of a row (column, respectively) of A to another row (column, respectively).

These are called the *elementary row (column, respectively) operations*. An *elementary matrix* is an $n \times n$ square matrix E which is obtained by performing exactly one of the three elementary operations on the $n \times n$ identity matrix I_n.

> **Theorems 3.1, 3.2.** Let $A \in \mathsf{M}_{m \times n}(F)$, and suppose that B is obtained from A by performing an elementary operation.
>
> - If this was a row operation, then $B = E\,A$ for some elementary $m \times m$ matrix E. Conversely, if E is an elementary $m \times m$ matrix A, then $E\,A$ is that matrix obtained from A by performing the same elementary row operation which produces E from I_m.
> - If this was a column operation, then $B = A\,E$ for some elementary $n \times n$ matrix E. Conversely, if E is an elementary $n \times n$ matrix A, then $A\,E$ is that matrix obtained from A by performing the same elementary column operation which produces E from I_n.
> - Elementary matrices are invertible, and the inverse of an elementary matrix is of the same type.

§3.2: The Rank of a Matrix and Matrix Inverses. Let $\mathsf{V} = \mathsf{F}^n$ and $\mathsf{W} = \mathsf{F}^m$, and denote β and γ as the standard basis for these vector spaces. Given any $m \times n$ matrix A, we define the linear transformation $\mathsf{L}_A : \mathsf{F}^n \to \mathsf{F}^m$ as $\mathsf{L}_A(x) = A\,x$. Rather explicitly,

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \implies \mathsf{L}_A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} + x_2 \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix}.$$

This linear transformation is called the *left-multiplication transformation*. Observe that $A = \left[\mathsf{L}_A\right]_\beta^\gamma$. We define the *rank* of the $m \times n$ matrix A as $\operatorname{rank}(A) = \operatorname{rank}(\mathsf{L}_A)$.

> **Theorems 3.3, 3.4.** Let A be an $m \times n$ matrix.
>
> - A is invertible if and only if both $m = n$ and $\operatorname{rank}(A) = n$.
> - Say that $A = \left[\mathsf{T}\right]_\beta^\gamma$ for some linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ between finite-dimensional vector spaces having bases β and γ. Then $\operatorname{rank}(\mathsf{T}) = \operatorname{rank}(A)$.
> - Denote $B = P\,A\,Q$ for some invertible $m \times m$ matrix P and invertible $n \times n$ matrix Q. Then $\operatorname{rank}(A) = \operatorname{rank}(B)$. In particular,
>   - Similar matrices have the same rank.
>   - Elementary operations on a matrix are rank preserving.

Theorem 3.6. Let A be an $m \times n$ matrix of rank r.

- $r \leq m$ and $r \leq n$.
- There exist invertible $m \times m$ and $n \times n$ matrices B and C, respectively, so that
$$D = B\,A\,C = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$
where I_r is the $r \times r$ identity matrix; while O_1 is an $r \times (n-r)$ matrix, O_2 is an $(m-r) \times r$ matrix, and O_3 is an $(m-r) \times (n-r)$ matrix consisting of all 0's.
- Both A and its transpose A^t have the same rank, that is, $\operatorname{rank}(A^t) = \operatorname{rank}(A)$.

Theorem 3.7.

- Let A be an $m \times n$ matrix and B be an $p \times n$ matrix. Then $\operatorname{rank}(B\,A) \leq \operatorname{rank}(A)$ and $\operatorname{rank}(B\,A) \leq \operatorname{rank}(B)$.
- Let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ and $\mathsf{U} : \mathsf{W} \to \mathsf{Z}$ be linear transformations between finite dimensional vector spaces. Then $\operatorname{rank}(\mathsf{U} \circ \mathsf{T}) \leq \operatorname{rank}(\mathsf{T})$ and $\operatorname{rank}(\mathsf{U} \circ \mathsf{T}) \leq \operatorname{rank}(\mathsf{U})$.

§3.3: Systems of Linear Equations – Theoretical Aspects. Say that we have a system of m linear equations in n unknowns:

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

We can write (S) in the single equation $A\,x = b$ in we invoke the matrices

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{mn} & \cdots & A_{mn} \end{pmatrix} \in \mathsf{M}_{m \times n}(F), \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathsf{F}^n, \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathsf{F}^m.$$

We make a few definitions about the system of linear equations (S).

- A vector $s \in \mathsf{F}^n$ is said to be a *solution* to the system (S) if $A\,s = b$. The collection of all solutions $s \in \mathsf{F}^n$ is called the *solution set* to the system (S).
- We say that (S) is *consistent* if there exists a vector $s \in \mathsf{F}^n$ such that $A\,s = b$. Otherwise, we say that (S) is *inconsistent* if no such vector s exists.
- We say that (S) is *homogeneous* if $b = O$ is the zero vector in F^m. Otherwise, we say that (S) is *inhomogeneous* if $b \neq O$.

Theorems 3.8, 3.9, 3.10. Say that (S) is a system of linear equations $A\,x = b$ consisting of m linear equations in n unknowns. Denote $\mathsf{V} = \mathsf{F}^n$ and $\mathsf{W} = \mathsf{F}^m$.

- Let $\mathsf{K_H} = \left\{ s \in \mathsf{V} \,\middle|\, A\,s = O_{\mathsf{W}} \right\}$ denote the set of solutions s to the honogemeous system $A\,x = O_{\mathsf{W}}$. Then $\mathsf{K_H} = \mathsf{N}(\mathsf{L}_A)$ is the null space for the linear transformation $\mathsf{L}_A : \mathsf{V} \to \mathsf{W}$. In particular, $\mathsf{K_H}$ is a subspace of V with $\dim(\mathsf{K_H}) = n - \operatorname{rank}(A)$.
- Say that $m < n$. Then there is a nonzero solution $s \neq O_{\mathsf{V}}$ to the homogeneous system $A\,x = O_{\mathsf{W}}$.
- Let $\mathsf{K} = \left\{ s \in \mathsf{V} \,\middle|\, A\,s = b \right\}$ denote the set of solutions s to the system (S). Say that (S) is consistent, and denote s_0 as one solution. Then all solutions to (S) are in the form $s = s_0 + k$ for some $k \in \mathsf{K_H}$. That is,
$$\mathsf{K} = \{s_0\} + \mathsf{K_H} = \left\{ s_0 + k \in \mathsf{V} \,\middle|\, k \in \mathsf{K_H} \right\}.$$
- Say that $m = n$. The system (S) has exactly one solution $s_0 \in \mathsf{V}$ if and only if A is an invertible $n \times n$ matrix.

§3.4: Systems of Linear Equations – Computational Aspects. A matrix is in *reduced row-echelon form (rref)* if it satisfies all of the following conditions:

(a) If a row has nonzero entries, then the first nonzero entry is 1, called the *pivot* in this row. The variables corresponding to the pivots are called *leading variables*, and the other variables are called *free variables*.
(b) If a column contains a pivot, then all other entries in that column are zero.
(c) If a row contains a pivot, then each row above contains a pivot further to the left.

Given a matrix A, here is a series of steps to arrive at a matrix B that is in reduced row-echelon form:

- *Step 1:* If the cursor entry is zero, swap the row with a row below having a nonzero entry in that column. This is a Type 1 operation.
- *Step 2:* Divide the cursor row with the nonzero entry to make the nonzero entry equal to 1. This is a Type 2 operation.
- *Step 3:* Eliminate all other entries in the cursor column by subtracting suitable multiples of the cursor row. This is a Type 3 operation.
- *Step 4:* Move the cursor down one row, and over to the right. If all the entries below are zero, continue to more to the right.

We denote $B = \operatorname{rref}(A)$ as its reduced row echelon form. This algorithm is known as *Gaussian Elimination.*

Theorem 3.16. Let A be an $m \times n$ matrix. Denote $B = \operatorname{rref}(A)$ as its reduced row echelon form, and let S denote the collection of pivot columns of B.

- S is a basis for $\mathsf{R}(\mathsf{L}_A) = \mathsf{R}(\mathsf{L}_B)$.
- $\operatorname{rank}(A) = \operatorname{rank}(B) = |S|$ is the number of nonzero rows of B.
- Say that $m = n$. Then A is invertible if and only if $\operatorname{rref}(A) = I_n$ is the $n \times n$ identity matrix.

§4.1: Determinants of Order 2.

Theorem 4.2. Let A be a 2×2 matrix. Then A is invertible if and only if $\det(A) \neq 0$.

Proposition.

- Let $S = \{u, v\}$ be a collection of two vectors in $\mathbb{R}^2$, and let $A \in \mathsf{M}_{2\times 2}(\mathbb{R})$ be that matrix having u and v as its columns. Then the parallelogram determined by S, namely the set
$$\mathcal{P} = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} x = a\,u + b\,v \text{ for some real numbers} \\ \text{satisfying } 0 \leq a \leq 1 \text{ and } 0 \leq b \leq 1 \end{array} \right\},$$
has area $\big|\det(A)\big|$. In particular, $\det(A) = 0$ if and only if its columns u and v are parallel.
- For any 2×2 matrix A, we have the identity
$$\big(A_{11}\,A_{12} + A_{21}\,A_{22}\big)^2 + \big(A_{11}\,A_{22} - A_{12}\,A_{21}\big)^2 = \big(A_{11}^2 + A_{21}^2\big)\,\big(A_{12}^2 + A_{22}^2\big).$$

The second statement is often called Lagrange's Identity.

§4.2: Determinants of Order n. Let A be an $n \times n$ matrix. Denote $\tilde{A}_{ij}$ as that $(n-1)\times(n-1)$ matrix obtained from A by deleting the ith row and jth column. We define the determinant of an $n \times n$ matrix A recursively as follows:

- If $n = 1$, then $\det(A) = A_{11}$.
- If $n \geq 2$, then $\det(A) = \displaystyle\sum_{j=1}^{n} (-1)^{1+j}\,A_{1j} \cdot \det(\tilde{A}_{1j})$.

The $n \times n$ matrix $\operatorname{cof}(A)$, the cofactor of A, is that matrix with entries

$$\big(\operatorname{cof}(A)\big)_{ij} = (-1)^{i+j}\,\det(\tilde{A}_{ij}) \qquad \text{where} \qquad \begin{array}{l} 1 \leq i \leq n, \\ 1 \leq j \leq n. \end{array}$$

The $n \times n$ matrix $\operatorname{adj}(A)$, the adjugate of A or the , is that matrix with entries

$$\big(\operatorname{adj}(A)\big)_{ij} = (-1)^{i+j}\,\det(\tilde{A}_{ji}) \qquad \text{where} \qquad \begin{array}{l} 1 \leq i \leq n, \\ 1 \leq j \leq n. \end{array}$$

Observe that the cofactor and classical adjoint are transposes of each other:

$$\operatorname{adj}(A) = \operatorname{cof}(A)^T.$$

Here are some examples.

- Let $n = 2$. Let's cross out the row and column where A_{11} and A_{12} appear:

$$\begin{array}{cc} A_{11} & \cancel{A_{12}} \\ \cancel{A_{21}} & A_{22} \end{array} \qquad\qquad \begin{array}{cc} \cancel{A_{11}} & A_{12} \\ A_{21} & \cancel{A_{22}} \end{array}$$

  This diagram says "$A_{11} \cdot \left|A_{22}\right|$" and "$A_{12} \cdot \left|A_{21}\right|$", respectively. We take the alternating sum of these two in order to compute the determinant:

$$\begin{aligned} \det(A) &= (-1)^{1+1} A_{11} \det(\tilde{A}_{11}) + (-1)^{1+2} A_{12} \det(\tilde{A}_{12}) \\ &= A_{11} A_{22} - A_{12} A_{21}. \end{aligned}$$

  The cofactor matrix is

$$\operatorname{cof}(A) = \begin{pmatrix} (-1)^{1+1} \det(\tilde{A}_{11}) & (-1)^{1+2} \det(\tilde{A}_{12}) \\ (-1)^{2+1} \det(\tilde{A}_{21}) & (-1)^{2+2} \det(\tilde{A}_{22}) \end{pmatrix} = \begin{pmatrix} A_{22} & -A_{21} \\ -A_{12} & A_{11} \end{pmatrix}.$$

  The classical adjoint is the transpose of this matrix, that is, $\operatorname{adj}(A) = \operatorname{cof}(A)^T$. If $\det(A) \neq 0$, then A has inverse

$$A^{-1} = \frac{1}{A_{11} A_{22} - A_{12} A_{21}} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

- Let $n = 3$. Let's cross out the row and column where A_{11}, A_{12} and A_{13} appear:

$$\begin{array}{ccc} A_{11} & \cancel{A_{12}} & \cancel{A_{13}} \\ \cancel{A_{21}} & A_{22} & A_{23} \\ \cancel{A_{31}} & A_{32} & A_{33} \end{array} \qquad \begin{array}{ccc} \cancel{A_{11}} & A_{12} & \cancel{A_{13}} \\ A_{21} & \cancel{A_{22}} & A_{23} \\ A_{31} & \cancel{A_{32}} & A_{33} \end{array} \qquad \begin{array}{ccc} \cancel{A_{11}} & \cancel{A_{12}} & A_{13} \\ A_{21} & A_{22} & \cancel{A_{23}} \\ A_{31} & A_{32} & \cancel{A_{33}} \end{array}$$

  This diagram says "$A_{11} \cdot \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix}$", "$A_{12} \cdot \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix}$", and "$A_{13} \cdot \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}$", respectively. We take the alternating sum of these three in order to compute the determinant:

$$\begin{aligned} \det(A) &= (-1)^{1+1} A_{11} \det(\tilde{A}_{11}) + (-1)^{1+2} A_{12} \det(\tilde{A}_{12}) + (-1)^{1+3} A_{13} \det(\tilde{A}_{13}) \\ &= A_{11} \cdot \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12} \cdot \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13} \cdot \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} \\ &= A_{11} (A_{22} A_{33} - A_{23} A_{32}) - A_{12} (A_{21} A_{33} - A_{23} A_{31}) + A_{13} (A_{21} A_{32} - A_{22} A_{31}) \\ &= (A_{11} A_{22} A_{33} + A_{12} A_{23} A_{31} + A_{13} A_{21} A_{32}) \\ &\qquad - (A_{11} A_{23} A_{32} + A_{12} A_{21} A_{33} + A_{13} A_{22} A_{31}). \end{aligned}$$

We can also express this formula using the following diagram:

$$
\begin{array}{ccccc}
A_{11} & A_{12} & A_{13} & A_{11} & A_{12} \\
A_{21} & A_{22} & A_{23} & A_{21} & A_{22} \\
A_{31} & A_{32} & A_{33} & A_{31} & A_{32}
\end{array}
$$

Recall that the arrows mean "multiply all of these numbers together", where right-ward arrows correspond to "+" and the left-ward arrows correspond to "−". The adjugate matrix is

$$
\operatorname{cof}(A) = \begin{pmatrix} (-1)^{1+1}\det(\tilde{A}_{11}) & (-1)^{1+2}\det(\tilde{A}_{12}) & (-1)^{1+3}\det(\tilde{A}_{13}) \\ (-1)^{2+1}\det(\tilde{A}_{21}) & (-1)^{2+2}\det(\tilde{A}_{22}) & (-1)^{2+3}\det(\tilde{A}_{23}) \\ (-1)^{3+1}\det(\tilde{A}_{31}) & (-1)^{3+2}\det(\tilde{A}_{32}) & (-1)^{3+3}\det(\tilde{A}_{33}) \end{pmatrix}
$$

$$
= \begin{pmatrix} +\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} & -\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} & +\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} \\ -\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} & +\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} & -\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} \\ +\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} & -\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix} & +\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \end{pmatrix}.
$$

The classical adjoint is the transpose of this matrix, that is, $\operatorname{adj}(A) = \operatorname{cof}(A)^T$. If $\det(A) \neq 0$, then A has inverse

$$
A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).
$$

use cofector to calculate A inverse

Theorem 4.3. Let $\{a_1, a_2, \ldots, a_n\}$ denote the rows of an $n \times n$ matrix A. For each integer r satisfying $1 \leq r \leq n$, define the function $\mathsf{T}_r : \mathsf{F}^n \to \mathsf{F}^1$ by $\mathsf{T}_r(x) = \det(B)$ where B is that $n \times n$ matrix obtained from A by replacing row r, namely a_r, with x:

$$
B_{ij} = \begin{cases} A_{ij} & \text{if } i \neq r, \\ x_j & \text{if } i = r; \end{cases} \qquad \text{where} \qquad \begin{array}{l} 1 \leq i \leq n, \\ 1 \leq j \leq n. \end{array}
$$

Then T_r is a linear transformation. That is, the determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed.

§4.3: Properties of Determinants.

Theorem 4.4. Let A be an $n \times n$ matrix.

- For any row i, we have the expression

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}).$$

That is, the determinant of a square matrix can be evaluated by cofactor expansion along any row.

- $\det(A) = 0$ whenever A has a row of zeroes.
- $\det(A) = 0$ whenever A has two identical rows.

cofactor get det(A)

det(A) = 0 的情况

Let $A \in \mathsf{M}_{m\times n}(F)$ be an $m \times n$ matrix, that is, an array with m rows and n columns:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \qquad \text{where} \qquad A_{ij} \in F.$$

There are three types of operations we can perform on the rows (columns, respectively) of A:

- *Type 1:* Interchanging any two rows (columns, respectively) of A.
- *Type 2:* Multiplying any row (column, respectively) of A by a nonzero scalar.
- *Type 3:* Adding a scalar multiple of a row (column, respectively) of A to another row (column, respectively).

These are called the *elementary row (column, respectively) operations.* An *elementary matrix* is an $n \times n$ square matrix E which is obtained by performing exactly one of the three elementary operations on the $n \times n$ identity matrix I_n.

Theorem 4.5, 4.6. Let A be an $n \times n$ square matrix.

det(A) self operation

- Let B be an $n \times n$ matrix obtained from A by interchanging two distinct rows. Then $\det(B) = -\det(A)$.
- Let B be an $n \times n$ matrix obtained from A multiplying one row of A by a (nonzero) scalar k. Then $\det(B) = k \det(A)$.
- Let B be an $n \times n$ matrix obtained from A adding a multiple of one row of A to another row of A. Then $\det(B) = +\det(A)$.

Corollary Let E be an $n \times n$ elementary matrix, and let I_n be the $n \times n$ identity matrix.

- Let E be an $n \times n$ matrix obtained from I_n by interchanging two distinct rows. Then $\det(E) = -1$.
- Let E be an $n \times n$ matrix obtained from I_n multiplying one row of I_n by a nonzero scalar k. Then $\det(E) = k$.
- Let E be an $n \times n$ matrix obtained from I_n adding a multiple of one row of I_n to another row of I_n. Then $\det(E) = 1$.

§4.4: Summary – Important Facts about Determinants.

det(AB) multiplication

Theorem 4.7. Let A be an $n \times n$ matrix.
- For any $n \times n$ matrix B, we have $\det(A\,B) = \det(A)\,\det(B)$.
- A is invertible if and only if $\det(A) \neq 0$. In particular, $\det(A^{-1}) = \det(A)^{-1}$.
- $\det(A^t) = \det(A)$.

det(a) != 0 <=> A invertible
inverse determinant property

Theorem 4.9 (Cramer's Rule). Say that we have a system of n linear equations in n unknowns:

$$(S) \qquad \begin{array}{ccccccccc} A_{11}\,x_1 & + & A_{12}\,x_2 & + & \cdots & + & A_{1n}\,x_n & = & b_1 \\ A_{21}\,x_1 & + & A_{22}\,x_2 & + & \cdots & + & A_{2n}\,x_n & = & b_2 \\ & & & & & & & \vdots & \\ A_{n1}\,x_1 & + & A_{n2}\,x_2 & + & \cdots & + & A_{mn}\,x_n & = & b_m \end{array}$$

which we write as the single equation $A\,x = b$ in terms of the matrices

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

det(A) != 0
adjugate = transpose of cof

Assume that $\det(A) \neq 0$.
- $A^{-1} = \dfrac{\operatorname{adj}(A)}{\det(A)}$ in terms of the classical adjoint of A.
- The system (S) has a unique solution $s_0 \in \mathsf{F}^n$, namely
$$s_0 = \frac{\operatorname{adj}(A)\,b}{\det(A)}, \qquad \text{that is,} \qquad x_k = \frac{\det(M_k)}{\det(A)} \quad \text{for } 1 \leq k \leq n$$
where M_k is the $n \times n$ matrix obtained from A by replacing column k by b.

1. get inverse
2. get solution

Here is an example when $n = 2$. Consider a system of linear equations in the form

$$(S) \qquad \begin{array}{ccccc} A_{11}\,x_1 & + & A_{12}\,x_2 & = & b_1 \\ A_{21}\,x_1 & + & A_{22}\,x_2 & = & b_2 \end{array}$$

We can express the system of linear equations (S) in the form $A\,x = b$ in terms of the matrices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \text{and} \qquad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Assuming that $\det(A) = A_{11}\,A_{22} - A_{12}\,A_{21} \neq 0$, we can find the desired solution as

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A^{-1}\,b = \frac{1}{A_{11}\,A_{22} - A_{12}\,A_{21}} \begin{pmatrix} +A_{22}\,b_1 - A_{12}\,b_2 \\ -A_{21}\,b_1 + A_{11}\,b_2 \end{pmatrix}.$$

in other words,

$$x_1 = \frac{\det\begin{pmatrix} b_1 & A_{12} \\ b_2 & A_{22}\end{pmatrix}}{\det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}} \qquad \text{and} \qquad x_2 = \frac{\det\begin{pmatrix} A_{11} & b_1 \\ A_{21} & b_2\end{pmatrix}}{\det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}}.$$

§5.1: Eigenvalues and Eigenvectors.

A is a diagonal matrix

- We say that T is *diagonalizable* if there is an ordered basis β such that $[\mathsf{T}]_\beta^\beta$ is a a diagonal matrix, that is, if we can write

T is linear transformation of Vj

$$\text{(Aij) = matrix A = } [\mathsf{T}]_\beta^\beta = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \qquad \Longleftrightarrow \qquad \mathsf{T}(v_j) = \lambda_j\, v_j \text{ for } 1 \le j \le n. \text{ = sum(Aij*vi)}$$

- A square matrix $A \in \mathsf{M}_{n\times n}(F)$ is said to be *diagonalizable* if the linear transformation "left multiplication by A," namely $\mathsf{L}_A : \mathsf{F}^n \to \mathsf{F}^n$ defined by $x \mapsto A\,x$, is diagonalizable. To be more precise, let $\gamma = \{e_1, e_2, \ldots, e_n\}$ denote the standard basis of $\mathsf{V} = \mathsf{F}^n$; Theorem 2.15 asserts that, $A = [\mathsf{L}_A]_\gamma^\gamma$. For any other basis $\beta = \{v_1, v_2, \ldots, v_n\}$, Theorem 2.11 asserts that

$$\begin{aligned} [\mathsf{L}_A]_\gamma^\gamma\, [\mathsf{I}_\mathsf{V}]_\beta^\gamma &= [\mathsf{L}_A \circ \mathsf{I}_\mathsf{V}]_\beta^\gamma \\ &= [\mathsf{L}_A]_\beta^\gamma \\ &= [\mathsf{I}_\mathsf{V} \circ \mathsf{L}_A]_\beta^\gamma \\ &= [\mathsf{I}_\mathsf{V}]_\beta^\gamma\, [\mathsf{L}_A]_\beta^\beta \end{aligned} \qquad \Longrightarrow \qquad Q^{-1}\, A\, Q = [\mathsf{L}_A]_\beta^\beta = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

where $Q = [\mathsf{I}_\mathsf{V}]_\beta^\gamma$ is that invertible $n \times n$ matrix whose column j is the vector v_j.

dnf: eigenvalue, eigenvector

- Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation on a finite dimensional vector space V. A nonzero vector $v \in \mathsf{V}$ is called an *eigenvector* of T if there exists a scalar $\lambda \in F$ such that $\mathsf{T}(v) = \lambda\, v$. The scalar λ is called the *eigenvalue* associated to the eigenvector v.

- Let $A \in \mathsf{M}_{n\times n}(F)$. A vector $x \in \mathsf{F}^n$ is called an *eigenvector* of A if there exists a scalar $\lambda \in F$ such that $A\,x = \lambda\,x$. The scalar λ is called the *eigenvalue* associated to the eigenvector x.

eigenvalue, vectors property

Theorem 5.1, 5.2, 5.3, 5.4. Let V be a finite dimensional vector space over a field F with basis β. Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation, and denote $A = \left[\mathsf{T}\right]_\beta^\beta$ as its matrix representation.

- T is diagonalizable if and only if there exists an ordered basis $\beta = \{v_1, v_2, \ldots, v_n\}$ consisting of eigenvectors v_j of T. Furthermore, if T is diagonalizable, then A is a diagonal matrix and $A_{jj} = \lambda_j$ is the eigenvalue associated to v_j for $1 \le j \le n$.

A diagonal matrix:

- The following are equivalent for $\lambda \in F$:
  i. λ is an eigenvalue for T.
  ii. λ is an eigenvalue for A.
  iii. $\det\big(A - \lambda\, I_n\big) = 0$. = characteristic polynomial of T
  iv. There exists $v \neq O_{\mathsf{V}}$ such that $v \in \mathsf{E}_\lambda = \mathsf{N}\big(\mathsf{T} - \lambda\, \mathsf{I}_{\mathsf{V}}\big)$.
- For any variable t, the $n \times n$ determinant
$$f(t) = \det\big(A - t\, I_n\big) = \det(A) + \cdots + (-1)^n\, t^n$$
is a polynomial of degree n which is independent of choice of basis β. In particular, T has at most n distinct eigenvalues λ.

We use this result to compute the eigenvalues λ of a linear transformation $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ as follows.

#1. Choose a basis $\beta = \{v_1, v_2, \ldots, v_n\}$ of V.
#2. Compute the $n \times n$ matrix $A = \left[\mathsf{T}\right]_\beta^\beta$.
#3. Compute the degree n polynomial $f(t) = \det\big(A - t\, I_n\big) = \det(A) + \cdots + (-1)^n\, t^n$.
#4. Compute the roots λ of the equation $f(\lambda) = 0$. These roots λ are the eigenvalues of T.

The polynomial $f(t)$ is called the *characteristic polynomial of T*.

§5.2: Diagonalizability.

Theorem 5.5. Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation for an n-dimensional vector space V.

- Assume that T has $k \le n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. If $S = \{v_1, v_2, \ldots, v_k\}$ is a collection of nonzero vectors such that $\mathsf{T}(v_j) = \lambda_j\, v_j$ for $1 \le j \le k$, then S is a linearly independent set.
- If T has n distinct eigenvalues, then T is diagonalizable.

- If T has n distinct eigenvalues, then a collection $\beta = \{v_1, v_2, \ldots, v_n\}$ of eigenvectors is a basis. For example, the reflection $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ about the line $y = m\, x$ in $\mathsf{V} = \mathbb{R}^2$ has $n = 2$ distinct eigenvalues, namely $\lambda_1 = +1$ and $\lambda_2 = -1$. We have seen that its eigenvectors
$$v_1 = \frac{1}{\sqrt{1+m^2}} \begin{pmatrix} 1 \\ m \end{pmatrix} \quad \text{and} \quad v_2 = \frac{1}{\sqrt{1+m^2}} \begin{pmatrix} -m \\ 1 \end{pmatrix} \quad \Longrightarrow \quad \left[\mathsf{T}\right]_\beta^\beta = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}$$
form a basis $\beta = \{v_1, v_2\}$ for V.

- If T has $k \neq n$ distinct eigenvalues, then T may still be diagonalizable – and its eigenvectors may still form a basis. For example, the identity map $\mathsf{I}_{\mathsf{V}} : \mathsf{V} \to \mathsf{V}$ on $\mathsf{V} = \mathsf{F}^2$ has just $k = 1$

eigenvalue, namely $\lambda_1 = 1$. We have seen that its eigenvectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \Longrightarrow \qquad \big[\mathsf{I_V}\big]_\gamma^\gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

form the standard basis $\gamma = \{e_1, e_2\}$ for V.

Let V be a n-dimensional vector space over a field F with basis β. Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation, and denote $A = \big[\mathsf{T}\big]_\beta^\beta$ as its matrix representation. Assume that T has $k \leq n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then we have the characteristic polynomial

$$f(t) = \det\big(A - t\, I_n\big) = \det(A) + \cdots + (-1)^n\, t^n = (-1)^n \, \big(t - \lambda_1\big)^{m_1} \, \big(t - \lambda_2\big)^{m_2} \cdots \big(t - \lambda_k\big)^{m_k}$$

for some exponents $m_1, m_2, \ldots, m_k$. We say that m_j is the *(algebraic) multiplicity* of λ_j. Observe that the sum of the multiplicities must be the dimension of the vector space:

$$m_1 + m_2 + \cdots + m_k = \deg(f) = n = \dim(\mathsf{V}).$$

Recall that Theorem 5.4 asserts that λ is an eigenvalue of T if and only if there exists $v \neq O_\mathsf{V}$ contained in the null space

$$\text{eigenspace} = \ \mathsf{E}_\lambda = \mathsf{N}\big(\mathsf{T} - \lambda\, \mathsf{I_V}\big) = \left\{ v \in \mathsf{V} \,\middle|\, \mathsf{T}(v) = \lambda\, v \right\}.$$

We call this subspace the *eigenspace of T corresponding to* λ.

distinct eigenvalues property

Theorem 5.7, 5.8, 5.9. Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation for an n-dimensional vector space V. Assume that T has $k \leq n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, where each λ_i has multiplicity m_i.

- $1 \leq \dim(\mathsf{E}_{\lambda_i}) \leq m_i$.
- $\mathsf{E}_{\lambda_i} \cap \mathsf{E}_{\lambda_j} = \{O_\mathsf{V}\}$ whenever $i \neq j$.
- Let S_i be a linearly independent subset of E_{λ_i}. Then the union $S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a linearly independent subset of V.
- T is diagonalizable if and only if $\dim(\mathsf{E}_{\lambda_i}) = m_i$ for $1 \leq i \leq k$. If T is indeed diagonalizable, let $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ denote the union of bases β_i for E_{λ_i}. Then $\beta = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis for V consisting of eigenvectors v_j for T.

distinct eigenvalues with diagonal

§5.3: Matrix Limits. Denote $F = \mathbb{C}$ as the collection of complex numbers. Following Appendix D in the text, denote the *absolute value* or the *modulus* of a complex number $z = a + b\, i$ as the real number $|z| = \sqrt{a^2 + b^2}$. Fix an extended number p satisfying $1 \leq p \leq \infty$. For any n-dimensional vector define the *p-norm* as the real number

$$\|x\|_p = \begin{cases} \displaystyle\sum_{j=1}^n |x_j| & \text{when } p = 1, \\ \left(\displaystyle\sum_{j=1}^n |x_j|^p \right)^{1/p} & \text{when } 1 < p < \infty, \text{ and} \\ \max\Big\{ |x_1|, |x_2|, \ldots, |x_n| \Big\} & \text{when } p = \infty; \end{cases} \qquad \text{where} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

When $p = 1$ this is called the *city block distance*, the *Manhattan distance*, or the *Manhattan length*; when $p = 2$, this is called the *Euclidean norm*; and when $p = \infty$ this is called the *maximum norm*, the *uniform norm*, or the *Chebyshev distance*.

Proposition. Fix two extended numbers p and q satisfying

$$1 \le p \le \infty, \qquad 1 \le q \le \infty, \qquad \text{and} \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

For x, $y \in \mathbb{C}^n$ and $c \in \mathbb{C}$, the p-norm satisfies the following properties.

- *Positivity:* $\|x\|_p \ge 0$.
- *Non-degeneracy:* $\|x\|_p = 0$ if and only if x is the zero vector.
- *Multiplicativity:* $\|c\,x\|_p = |c|\,\|x\|_p$.
- *Hölder's Inequality:* $\sum_{j=1}^n |x_j\,y_j| \le \|x\|_p\,\|y\|_q$.
- *Minkowski's Inequality:* $\|x+y\|_p \le \|x\|_p + \|y\|_p$.

Note that when $p = 1$ we have $q = \infty$; and when $p = 2$ we have $q = 2$. When $p = q = 2$, Hölder's Inequality is the same as the *Cauchy-Bunyakovsky-Schwarz (CBS) Inequality*, and Minkowski's Inequality is the same as the *Triangle Inequality*:

$$\left|\sum_{j=1}^n x_j\,\overline{y_j}\right| \le \sqrt{\sum_{j=1}^n |x_j|^2} \cdot \sqrt{\sum_{j=1}^n |y_j|^2} \qquad \text{and} \qquad \sqrt{\sum_{j=1}^n |x_j+y_j|^2} \le \sqrt{\sum_{j=1}^n |x_j|^2} + \sqrt{\sum_{j=1}^n |y_j|^2}.$$

Now that we have the concept of a p-norm on vectors $x \in \mathbb{C}^n$, we can extend this to the concept of a p-norm on matrices. For any $m \times n$ matrix A in $\mathsf{M}_{m\times n}(\mathbb{C})$, define the *p-norm* or the *induced norm* or the *operator norm* as the real number

$$\|A\|_p = \sup\left\{ \frac{\|A\,x\|_p}{\|x\|_p} \;\middle|\; x \in \mathbb{C}^n,\ x \ne (0,0,\dots,0) \right\}.$$

If $x \in \mathbb{C}^n$ is not the zero vector, then we found above that $\|x\|_p \ne 0$, so that the ratio $\|A\,x\|_p/\|x\|_p$ is a well-defined nonnegative real number. The quantity $\|A\|_p$ is the supremum over all such ratios: you can think of this p-norm as the largest value that the ratio $\|A\,x\|_p/\|x\|_p$ can get. Observe by definition that

$$\|A\,x\|_p \le \|A\|_p\,\|x\|_p \qquad \text{for all } x \in \mathbb{C}^n.$$

This norm has properties similar to those above.

Proposition. Fix an extended number p satisfying $1 \leq p \leq \infty$. For $A,\, B \in \mathsf{M}_{m\times n}(\mathbb{C})$ and $c \in \mathbb{C}$, the p-norm satisfies the following properties.

- *Positivity:* $\|A\|_p \geq 0$.
- *Non-degeneracy:* $\|A\|_p = 0$ if and only if A is the $m \times n$ zero matrix.
- *Multiplicativity:* $\|c\,A\|_p = |c|\;\|A\|_p$.
- *Minkowski's Inequality:* $\|A+B\|_p \leq \|A\|_p + \|B\|_p$.
- We have the values

$$\rho(A) = \max\left\{ \rho_i(A) \,\middle|\, 1 \leq i \leq m \right\} = \|A\|_\infty, \qquad \rho_i(A) = \sum_{j=1}^{n} |A_{ij}|;$$

$$\nu(A) = \max\left\{ \nu_j(A) \,\middle|\, 1 \leq j \leq n \right\} = \|A\|_1, \qquad \nu_j(A) = \sum_{i=1}^{m} |A_{ij}|.$$

- If $\lambda \in \mathbb{C}$ is an eigenvalue of A, then

$$|\lambda| \leq \inf\left\{ \|A\|_p \,\middle|\, 1 \leq p \leq \infty \right\} \leq \min\bigl\{\rho(A),\, \nu(A)\bigr\}.$$

Observe that $\rho_i(A)$ is the sum over the entries in row i, while $\nu_j(A)$ is the sum over the entries in column j. We call $\rho(A)$ the *row sum of* A, and $\nu(A)$ the *column sum of* A. Note that they are special cases of $\|A\|_p$ of the operator norm for $p = \infty$ and $p = 1$, respectively.
Let $A \in \mathsf{M}_{n\times n}(\mathbb{C})$ be an $n \times n$ square matrix. Recall that an eigenvalue λ is a root of the characteristic polynomial

$$f(t) = \det\bigl(A - t\,I_n\bigr) = \det(A) + \cdots + (-1)^n\, t^n.$$

We wish to give an algorithm to approximate these eigenvalues using the row sum above.

Theorem 5.16 (Gerschgorin's Disk Theorem). Let $A \in \mathsf{M}_{n\times n}(\mathbb{C})$. For each row i of A, define the region

$$C_i = \left\{ z \in \mathbb{C} \,\middle|\, \bigl|z - A_{ii}\bigr| \leq \rho_i(A) - \bigl|A_{ii}\bigr| \right\} \qquad \text{where} \qquad \rho_i(A) = \sum_{j=1}^{n} |A_{ij}|$$

as a disk of radius $\rho_i(A) - \bigl|A_{ii}\bigr|$ centered around the ith diagonal entry A_{ii}. Then each eigenvalue $\lambda \in C_i$ for some $1 \leq i \leq n$.

The C_i's are sometimes called *Gerschgorin Disks.* Observe that if $0 \notin C_i$ for $1 \leq i \leq n$, then each eigenvalue $\lambda \neq 0$. Since the characteristic polynomial factors as

$$\begin{aligned} f(t) &= \det\bigl(A - t\,I_n\bigr) \\ &= (-1)^n\,\bigl(t-\lambda_1\bigr)^{m_1}\,\bigl(t-\lambda_2\bigr)^{m_2}\cdots\bigl(t-\lambda_k\bigr)^{m_k} \end{aligned} \qquad \implies \qquad \det(A) = f(0) = \lambda_1^{m_1}\,\lambda_2^{m_2}\cdots\lambda_k^{m_k} \neq 0.$$

The Corollary to Theorem 4.7 asserts that A is an invertible matrix. Hence Gerschgorin Disks give a computational way to determine whether a matrix is invertible.

Consider a sequence $\{z_1,\, z_2,\, \ldots,\, z_k,\, \ldots\} \subseteq \mathbb{C}$ of complex numbers. The expression

$$\lim_{k\to\infty} z_k = z_\infty$$

for some complex number z_∞ means, given a real number $\epsilon > 0$, we can find a real number $\delta > 0$ such that $|z_k - z_\infty| < \epsilon$ whenever $k > \delta$. We also say that the sequence "converges to a limit ℓ".

Similarly, the expression

$$\lim_{k\to\infty} z_k = \infty$$

means, given a real number $\epsilon > 0$, we can find a real number $\delta > 0$ such that $|z_k| > \epsilon$ whenever $k > \delta$. We also say that the sequence "increases without bound."

Proposition. Let $\{ A_1, A_2, \ldots, A_k, \ldots, \} \subseteq \mathsf{M}_{m\times n}(\mathbb{C})$ be a sequence of $m \times n$ matrices with complex entries. Then the following are equivalent for an $m \times n$ matrix L:

i. For *each* extended number p satisfying $1 \leq p \leq \infty$, we can find a $\delta > 0$ such that for any given $\epsilon > 0$ we have $\|A_k - L\|_p < \epsilon$ whenever $k > \delta$.
ii. For *some* extended number p_0 satisfying $1 \leq p_0 \leq \infty$, we can find a $\delta > 0$ such that for any given $\epsilon > 0$ we have $\|A_k - L\|_{p_0} < \epsilon$ whenever $k > \delta$.
iii. We have the $m\,n$ limits

$$\lim_{k\to\infty} (A_k)_{ij} = L_{ij} \qquad \text{for all} \qquad \begin{matrix} 1 \leq i \leq m, \\ 1 \leq j \leq n. \end{matrix}$$

each entry has a limit

If any of these equivalent statements is true, then we write "$\lim_{k\to\infty} A_k = L$" and we say that the sequence "converges to a limit L".

complex matrix limitation thm

Theorem 5.12. Let $\{A_1, A_2, \ldots, A_k, \ldots, \} \subseteq \mathsf{M}_{m\times n}(\mathbb{C})$ be a sequence of $m \times n$ matrices with complex entries which converges to a limit L.

- Let $c \in \mathbb{C}$. Then the sequence $\{\ldots, c\,A_k, \ldots\} \subseteq \mathsf{M}_{m\times n}(\mathbb{C})$ tends to the limit $c\,L$. lim (k->inf) c Ak = c Lk
- Let $P \in \mathsf{M}_{q\times m}(\mathbb{C})$. Then the sequence $\{\ldots, P\,A_k, \ldots\} \subseteq \mathsf{M}_{q\times n}(\mathbb{C})$ tends to the limit $P\,L$.
- Let $Q \in \mathsf{M}_{n\times q}(\mathbb{C})$. Then the sequence $\{\ldots, A_k\,Q, \ldots\} \subseteq \mathsf{M}_{m\times q}(\mathbb{C})$ tends to the limit $L\,Q$.
- Assume that each $A_k = A^k \in \mathsf{M}_{n\times n}(\mathbb{C})$ are powers of a square matrix, and let $Q \in \mathsf{M}_{n\times n}(\mathbb{C})$ be invertible. Then the sequence $\{\ldots, Q^{-1}\,A^k\,Q, \ldots\} \subseteq \mathsf{M}_{n\times n}(\mathbb{C})$ tends to the limit $Q^{-1}\,L\,Q$.

lim (k->inf) Q-1 Ak Q = Q-1 L Q

§6.1: Inner Products and Norms. Denote $\mathsf{V} = \mathbb{C}^n$. The *length* of a vector $x \in \mathsf{V}$ is the scalar

$$||x||_2 = \sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2}.$$

We use this to motivate the following: We define a map $\langle\cdot,\,\cdot\rangle : \mathsf{V}\times\mathsf{V}\to\mathbb{C}$ by

complex conjugate

$$\langle x,\, y\rangle = a_1\,\overline{b_1} + a_2\,\overline{b_2} + \cdots + a_n\,\overline{b_n} \qquad \Longrightarrow \qquad \langle x,\, x\rangle = |a_1|^2 + |a_2|^2 + \cdots + |a_n|^2 = \|x\|_2^2.$$

This is known as the *standard inner product.*

F = R or F = C, inner product: V x V -> F
inner product properties

Proposition. For $x,\, y,\, z \in \mathbb{C}^n$ and $c \in \mathbb{C}$, the standard inner product has the following properties.

- *Linearity:* $\langle c\,x + z,\ y\rangle = c\,\langle x,\, y\rangle + \langle z,\, y\rangle$.
- *Symmetry:* $\overline{\langle x,\, y\rangle} = \langle y,\, x\rangle$.
- *Positivity:* $\langle x,\, x\rangle \geq 0$.
- *Non-degeneracy:* $\langle x,\, x\rangle = 0$ if and only if x is the zero vector.
- *Lagrange's Identity:*

$$\left|\sum_{k=1}^{n} a_k\,\overline{b_k}\right|^2 + \left(\sum_{i<j} \left|a_i\,b_j - a_j\,b_i\right|^2\right) = \left(\sum_{k=1}^{n} |a_k|^2\right)\left(\sum_{k=1}^{n} |b_k|^2\right).$$

In particular, $\left|\langle x,\, y\rangle\right| \leq \|x\|_2\,\|y\|_2$.

We wish to extend the concept of the standard inner product on $\mathsf{V} = \mathbb{C}^n$ to something similar for $\mathsf{V} = \mathsf{M}_{m\times n}(\mathbb{C})$. In general, let V be a vector space over $F = \mathbb{C}$. We say that a map $\langle\cdot,\,\cdot\rangle : \mathsf{V}\times\mathsf{V} \to \mathbb{C}$ is an *inner product* if the following properties hold for all $x,\, y,\, z \in \mathsf{V}$ and $c \in \mathbb{C}$:

- *Linearity:* $\langle c\,x + z,\ y\rangle = c\,\langle x,\, y\rangle + \langle z,\, y\rangle$.
- *Symmetry:* $\overline{\langle x,\, y\rangle} = \langle y,\, x\rangle$.
- *Positivity:* $\langle x,\, x\rangle \geq 0$.
- *Non-degeneracy:* $\langle x,\, x\rangle = 0$ if and only if $x = O_{\mathsf{V}}$.

Here are some examples.

- Let $\mathsf{V} = \mathbb{C}^n$. We have seen that the standard inner product $\langle\cdot,\,\cdot\rangle : \mathsf{V}\times\mathsf{V} \to \mathbb{C}$ defined by

$$\langle x,\, y\rangle = a_1\,\overline{b_1} + a_2\,\overline{b_2} + \cdots + a_n\,\overline{b_n} \qquad \text{where} \qquad x = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

  is an example of an inner product. More generally, choose n positive real numbers $r_1,\, r_2,\, \ldots,\, r_n$, and consider the new map $\langle\cdot,\,\cdot\rangle' : \mathsf{V}\times\mathsf{V} \to \mathbb{C}$ defined by

$$\langle x,\, y\rangle' = r_1\,a_1\,\overline{b_1} + r_2\,a_2\,\overline{b_2} + \cdots + r_n\,a_n\,\overline{b_n}.$$

  This is also an inner product.

- Let $\mathsf{V} = \mathsf{M}_{m\times n}(\mathbb{C})$ denote the collection of $m \times n$ matrices. If $B \in \mathsf{V}$ is an $m \times n$ matrix, define its *adjoint* $B^* \in \mathsf{M}_{n\times m}(\mathbb{C})$ as that $n \times m$ matrix found by computing the transpose of the complex conjugate of B. That is,

$$B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{pmatrix} \qquad \Longrightarrow \qquad B^* = \begin{pmatrix} \overline{B_{11}} & \overline{B_{21}} & \cdots & \overline{B_{m1}} \\ \overline{B_{12}} & \overline{B_{22}} & \cdots & \overline{B_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{B_{1n}} & \overline{B_{2n}} & \cdots & \overline{B_{mn}} \end{pmatrix}.$$

Define the map $\mathsf{V} \times \mathsf{V} \to \mathbb{C}$ which takes a pair A and B of $m \times n$ matrices to the trace of the $n \times n$ matrix $B^* A$:

$$\mathrm{tr}\,(B^*\,A) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}\,\overline{B_{ij}}.$$

This called the *Frobenius inner product*. Note that this reduces to the standard inner product when either $m = 1$ or $n = 1$. More generally, choose $m\,n$ positive real numbers $r_{11}, r_{12}, \ldots, r_{mn}$, and consider the map $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to \mathbb{C}$ defined by

$$\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}\,A_{ij}\,\overline{B_{ij}}.$$

This is always an inner product.

Let V be a vector space over $F = \mathbb{C}$, and say that a map $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to \mathbb{C}$ is an inner product. Since $\langle x, x \rangle \geq 0$, we define $\|x\| = \sqrt{\langle x, x \rangle}$ as the *norm* of $x \in \mathsf{V}$.

Norm defn and property

Theorem 6.2. Let V be a vector space over $F = \mathbb{C}$, and say that a map $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to \mathbb{C}$ is an inner product. Then for all $x, y \in \mathsf{V}$ and $c \in \mathbb{C}$, the map $\|\cdot\| : \mathsf{V} \to \mathbb{R}$ defined by $\|x\| = \sqrt{\langle x, x \rangle}$ has the following properties.

- *Positivity:* $\|x\| \geq 0$.
- *Non-degeneracy:* $\|x\| = 0$ if and only if x is the zero vector.
- *Multiplicativity:* $\|c\,x\| = |c|\,\|x\|$.
- *Cauchy-Bunyakovsky-Schwarz (CBS) Inequality:* $|\langle x, y \rangle| \leq \|x\|\,\|y\|$.
- *Triangle Inequality:* $\|x + y\| \leq \|x\| + \|y\|$.

Any map $\|\cdot\| : \mathsf{V} \to \mathbb{R}$ with the properties is called a *norm*. Observe that if V has an inner product, then it has a norm – but not all norms come from inner products! Indeed, we have p-norms $\|\cdot\|_p$ on $\mathbb{C}^n$ for any extended number p satisfying $1 \leq p \leq \infty$, but it only comes from an inner product if $p = 2$. A vector space V with a norm is called a *Banach Space*, while a vector space with an inner product is called an *Inner Product Space*.

§6.2: The Gram-Schmidt Orthogonalization Process and Orthogonal Complements. Let V be an n-dimensional linear space over either $F = \mathbb{R}$ or $F = \mathbb{C}$. Fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$. We say that a subset $\beta = \{u_1, u_2, \ldots, u_n\}$ of V is an *orthonormal basis* if

- *Span:* Given any $x \in \mathsf{V}$, we can find scalars a_i such that $x = a_1\,u_1 + a_2\,u_2 + \cdots + a_n\,u_n$.
- *Linear Independence:* The only scalars $a_i \in F$ which satisfy $a_1\,u_1 + a_2\,u_2 + \cdots + a_n\,u_n = O_{\mathsf{V}}$ are $a_1 = a_2 = \cdots = a_n = 0$.
- *Orthogonality:* Any pair $u_i, u_j \in \beta$ has the inner product $\langle u_i, u_j \rangle = 1$ whenever $i \neq j$.
- *Normalized:* Each vector in $u_j \in \beta$ has norm $\|u_j\| = \sqrt{\langle u_j, u_j \rangle} = 1$.

Orthonormal property

Here are some examples.

- Let $\mathsf{V} = \mathsf{F}^2$, and choose $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$ as the standard inner product. Recall that

$$\gamma = \{e_1, e_2\} \qquad \text{in terms of} \qquad e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

is the standard basis for V. We give a much more general statement: let $a,\, b \in \mathbb{R}$ be any scalars not both zero, and denote the two vectors

$$v_1 = \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} a \\ b \end{pmatrix} \qquad \text{and} \qquad v_2 = \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} -b \\ a \end{pmatrix}.$$

(We recover e_1 and e_2 when $a = 1$ and $b = 0$.) Then $\beta = \{v_1,\, v_2\}$ is an orthonormal basis for V. In particular, γ is an orthonormal basis for V, and there are infinitely many orthonormal bases for V.

- Let $\mathsf{V} = \mathsf{F}^3$, and choose $\langle \cdot,\, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$ as the standard inner product. Recall that

$$\gamma = \{e_1,\, e_2\} \qquad \text{in terms of} \qquad e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is the standard basis for V. We give a much more general statement: let $a,\, b,\, c \in \mathbb{R}$ be any scalars such that a and b are not both zero, and denote the three vectors

$$\begin{aligned} v_1 &= \frac{1}{\sqrt{a^2+b^2+c^2}} \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \\ v_2 &= \frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} -b \\ a \\ 0 \end{pmatrix}, \\ v_3 &= \frac{1}{\sqrt{a^2+b^2}\sqrt{a^2+b^2+c^2}} \begin{pmatrix} -a\,c \\ -b\,c \\ a^2+b^2 \end{pmatrix}. \end{aligned}$$

(We recover e_1, e_2, and e_3 when $a = 1$, $b = 0$, and $c = 0$.) Then $\beta = \{v_1,\, v_2,\, v_3\}$ is an orthonormal basis for V. In particular (like above), γ is an orthonormal basis for V, and there are infinitely many orthonormal bases for V.

Any finite dimensional inner product space V has an orthonormal basis

Theorem 6.3, 6.4, 6.5. Let V be a vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$.

- Let $\mathsf{W} = \operatorname{span}(S)$ be the span of an linearly independent subset $S = \{w_1, w_2, \ldots, w_k\}$ of V. Define $S' = \{v_1, v_2, \ldots, v_k\}$ by

$$v_i = \begin{cases} w_1 & \text{for } i = 1, \\ w_i - \displaystyle\sum_{j=1}^{i-1} \frac{\langle w_i, v_j \rangle}{\|v_j\|^2} v_j & \text{for } 2 \leq i \leq k. \end{cases}$$

  Then S' is an orthogonal set of nonzero vectors such that $\mathsf{W} = \operatorname{span}(S')$.
- Let $\mathsf{W} = \operatorname{span}(S')$ be the span of an orthogonal subset $S' = \{v_1, v_2, \ldots, v_k\}$ of V consisting of nonzero vectors, that is, $\langle v_i, v_j \rangle = 0$ for $i \neq j$ and $v_i \neq O_{\mathsf{V}}$. Then S' is a linearly independent set. Moreover, each $y \in \mathsf{W}$ can uniquely expressed in the form

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i.$$

ui is unit length, 2, 3 the same
They are to normalize

- Let $\mathsf{W} = \operatorname{span}(S'')$ be the span of an orthonormal subset $S'' = \{u_1, u_2, \ldots, u_k\}$ of V. Then S'' is a basis for W. Moreover, each $y \in \mathsf{W}$ can uniquely expressed in the form

$$y = \sum_{i=1}^{k} \langle y, u_i \rangle u_i.$$

The act of turning a linearly independent set $S = \{w_1, w_2, \ldots, w_k\}$ into an orthogonal set $S' = \{v_1, v_2, \ldots, v_k\}$ without zero vectors is called the *Gram-Schmidt Orthogonalization Process*. In particular, if V is a finite-dimensional vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, then V has an orthonormal basis β with respect to any given an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$. The construction goes as follows:

#1. Choose a basis $\gamma = \{w_1, w_2, \ldots, w_n\}$ for V.

#2. Construct an orthogonal basis $\{v_1, v_2, \ldots, v_n\}$ for V via the Gram-Schmidt Orthogonalization Process:

$$v_i = \begin{cases} w_1 & \text{for } i = 1, \\ w_i - \displaystyle\sum_{j=1}^{i-1} \frac{\langle w_i, v_j \rangle}{\|v_j\|^2} v_j & \text{for } 2 \leq i \leq n. \end{cases}$$

#3. Construct an orthonormal basis $\beta = \{u_1, u_2, \ldots, u_n\}$ for V via Normalizing:

$$u_i = \frac{1}{\|v_i\|} v_i \qquad \text{for } 1 \leq i \leq n.$$

Let V be a vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$. Given a subset $S \subseteq \mathsf{V}$, we define the *orthogonal complement of S* as the subset

$$S^{\perp} = \left\{ x \in \mathsf{V} \;\middle|\; \langle x, y \rangle = 0 \text{ for all } y \in S \right\}.$$

The symbol "$S^{\perp}$" is read as "S perp(endicular)".

Theorem 6.6, 6.7. Let V be a vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$. Let W be a finite-dimensional subspace.

- $\mathsf{W}^\perp$ is a subspace of V.
- $\mathsf{W} \cap \mathsf{W}^\perp = \{O_\mathsf{V}\}$.
- For each $y \in \mathsf{V}$, there exist unique $u \in \mathsf{W}$ and $z \in \mathsf{W}^\perp$ such that $y = u + z$. Moreover, the vector u is the unique vector in W that is "closest" to y; that is, for any $x \in \mathsf{W}$ we have the inequality $\|y - x\| \geq \|y - u\|$, with equality if and only if $x = u$.
- If V is finite-dimensional, then $\dim(\mathsf{W}) + \dim(\mathsf{W}^\perp) = \dim(\mathsf{V})$.

There exists a linear transformation $\mathrm{proj}_\mathsf{W} : \mathsf{V} \to \mathsf{V}$ which sends $y \mapsto u$. It is easy to see that $\mathsf{R}\big(\mathrm{proj}_\mathsf{W}\big) = \mathsf{W}$ is the image of such a map, while $\mathsf{N}\big(\mathrm{proj}_\mathsf{W}\big) = \mathsf{W}^\perp$ is the null space of such a map. Observe that

$$\big\| y - x \big\| \geq \big\| y - \mathrm{proj}_\mathsf{W}(y) \big\| \qquad \text{for any } x \in \mathsf{W} \text{ and } y \in \mathsf{V}.$$

Here is an example. Let $\mathsf{V} = \mathbb{R}^2$ denote the real plane along with the standard inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to \mathbb{R}$. Recall that V has an orthonormal basis $\beta = \{u_1, u_2\}$ in terms of the vectors

$$u_1 = \frac{1}{\sqrt{a^2 + b^2}} \begin{pmatrix} a \\ b \end{pmatrix} \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{a^2 + b^2}} \begin{pmatrix} -b \\ a \end{pmatrix}.$$

Theorem 6.5 asserts that any vector $v \in \mathsf{V}$ can be expressed in the form $v = v^\perp + v^{||}$, where

$$v^\perp = \langle v, u_1 \rangle \, u_1 = \frac{a\,x + b\,y}{a^2 + b^2} \begin{pmatrix} a \\ b \end{pmatrix} \qquad \text{and} \qquad v^{||} = \langle v, u_2 \rangle \, u_2 = \frac{a\,y - b\,x}{a^2 + b^2} \begin{pmatrix} -b \\ a \end{pmatrix}.$$

We give a geometric interpretation of this expression. Consider the line

$$\mathsf{W} = \mathrm{span}\,\{u_2\} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; a\,x + b\,y = 0 \right\}$$

going through the origin. We have the perpendicular line

$$\mathsf{W}^\perp = \mathrm{span}\,\{u_1\} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; b\,x - a\,y = 0 \right\}.$$

We can use this to define a linear transformation $\mathrm{proj}_\mathsf{W} : \mathsf{V} \to \mathsf{V}$ as the map $\mathrm{proj}_\mathsf{W}(v) = v^{||}$ which is the orthogonal projection onto W. Observe that $\mathsf{R}\big(\mathrm{proj}_\mathsf{W}\big) = \mathsf{W}$ is the image of such a map, while $\mathsf{N}\big(\mathrm{proj}_\mathsf{W}\big) = \mathsf{W}^\perp$ is the null space of such a map.

§6.3: The Adjoint of a Linear Operator.

Theorem 6.8, 6.9. Let V be a vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$.

- For each $y \in \mathsf{V}$, let $\mathsf{g} : \mathsf{V} \to F$ be the map which sends $x \mapsto \langle x, y \rangle$. Then g is a linear transformation. Conversely, let $\mathsf{g} : \mathsf{V} \to F$ be a linear transformation. Then there exists a unique $y \in \mathsf{V}$ such that $\mathsf{g}(x) = \langle x, y \rangle$ for all $x \in \mathsf{V}$.
- Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear transformation. There exists a unique linear transformation $\mathsf{T}^* : \mathsf{V} \to \mathsf{V}$ such that $\langle \mathsf{T}(x), y \rangle = \langle x, \mathsf{T}^*(y) \rangle$ for all $x, y \in \mathsf{V}$.

In other words, linear transformations $\mathsf{g} : \mathsf{V} \to F$ are in one-to-one correspondence with vectors $y \in \mathsf{V}$. The linear transformation $\mathsf{T}^* : \mathsf{V} \to \mathsf{V}$ is called the *adjoint* of T. The symbol "T^*" is read as "T-star."

Conjugate transpose

Theorem 6.10, 6.11. Let V be a vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle \cdot, \cdot \rangle : \mathsf{V} \times \mathsf{V} \to F$. Let $\mathsf{T}, \mathsf{U} : \mathsf{V} \to \mathsf{V}$ be linear transformations and $c \in F$ be a scalar.

- Let β be an orthonormal basis for V. Then the adjoint of $A = \left[\mathsf{T}\right]_\beta^\beta$ is the matrix $A^* = \left[\mathsf{T}^*\right]_\beta^\beta$ as the complex conjugate of the transpose of A.
- We have the following three properties:
  - $\left(c\,\mathsf{T} + \mathsf{U}\right)^* = \overline{c}\,\mathsf{T}^* + \mathsf{U}^*$
  - $\left(\mathsf{T} \circ \mathsf{U}\right)^* = \mathsf{U}^* \circ \mathsf{T}^*$
  - $\left(\mathsf{T}^*\right)^* = \mathsf{T}$.

Given an $n \times n$ matrix $A \in \mathsf{M}_{n\times n}(F)$, the $n \times n$ matrix A^* which is the complex conjugate of the transpose of A is called the *adjoint* of A. This should not be confused with the *classical adjoint* of A: this is that $n \times n$ matrix $\operatorname{adj}(A)$, also called the *adjugate* of A, with entries

$$\big(\operatorname{adj}(A)\big)_{ij} = (-1)^{i+j}\,\det(\tilde{A}_{ji}) \qquad \text{for} \qquad \begin{array}{c} 1 \le i \le n, \\ 1 \le j \le n; \end{array}$$

where $\tilde{A}_{ij}$ is that $(n-1)\times(n-1)$ matrix obtained from A by deleting the ith row and jth column. For example when $n = 2$ we have the matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \text{implies} \qquad A^* = \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} \qquad \text{but} \qquad \operatorname{adj}(A) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

Consider the special case where $\mathsf{V} = \mathsf{F}^n$. The standard inner product on V can be expressed in terms of the adjoint as follows:

$$\begin{aligned} \langle x,\, y\rangle &= x_1\,\overline{y_1} + x_2\,\overline{y_2} + \cdots + x_n\,\overline{y_n} \\ &= y^*\,x \end{aligned} \qquad \text{for all} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathsf{V}.$$

It is easy to see that the standard basis $\beta = \{e_1, e_2, \ldots, e_n\}$ is an orthonormal basis with respect to the standard inner product on V. Observe that

$$\left[\mathsf{L}_A\right]_\beta^\beta = A \qquad \Longrightarrow \qquad \left[(\mathsf{L}_A)^*\right]_\beta^\beta = A^* = \left[\mathsf{L}_{A^*}\right]_\beta^\beta \qquad \Longrightarrow \qquad (\mathsf{L}_A)^* = \mathsf{L}_{A^*}.$$

Say that in an experiment we have the following data:

Time	t_1	t_2	t_3	$\cdots$	t_m
Value	y_1	y_2	y_3	$\cdots$	y_m

The "least squares best linear fit" $y(t) = c\,t + d$ has coefficients

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} t_i^2 & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} 1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{m} t_i\, y_i \\ \sum_{i=1}^{m} y_i \end{pmatrix}.$$

This is part of a general phenomenon.

> **Theorem 6.12.** Let $A \in \mathsf{M}_{m\times n}(F)$ and $y \in \mathsf{F}^n$ for either $F = \mathbb{R}$ or $F = \mathbb{C}$. The vector $x_0 = \big(A^*\,A\big)^{-1}\big(A^*\,y\big)$ in F^n satisfies $||A\,x_0 - y|| \leq ||A\,x - y||$ for all $x \in \mathsf{F}^n$.

The expression $\big(A^*\,A\big)\,x_0 = \big(A^*\,y\big)$ is called the *normal equation*.

§6.4: Self-Adjoint Operators. Let V be a finite dimensional vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle\cdot,\,\cdot\rangle : \mathsf{V}\times\mathsf{V} \to F$. A linear operator $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ is said to be *normal* if $\mathsf{T}\,\mathsf{T}^* = \mathsf{T}^*\,\mathsf{T}$.

> **Theorem 6.15.** Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a normal linear operator on a finite dimensional inner product space V.
>
> - $\|\mathsf{T}(x)\| = \|\mathsf{T}^*(x)\|$ for all $x \in \mathsf{V}$.
> - $\mathsf{T} - c\,\mathsf{I}_\mathsf{V}$ is normal for any scalar $c \in F$.
> - Say that $x \in \mathsf{V}$ is an eigenvector of T with eigenvalue $\lambda \in F$, that is, $\mathsf{T}(x) = \lambda\,x$. Then x is also an eigenvector of T^* but with eigenvalue the complex conjugate of λ, that is, $\mathsf{T}^*(x) = \overline{\lambda}\,x$.
> - Say that $x_1,\,x_2 \in \mathsf{V}$ are eigenvectors of T corresponding to distinct eigenvalues $\lambda_1,\,\lambda_2 \in F$. Then x_1 and x_2 are orthogonal.

> **Theorem 6.14, 6.16.** Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear operator on a n-dimensional inner product space V.
>
> - (Issai Schur) Assume that the characteristic polynomial of T splits over F, that is, the polynomial
>
> $$\begin{aligned} f(t) = \det\Big(\big[\mathsf{T}\big]_\beta^\beta - t\,I_n\Big) &= \det(\mathsf{T}) + \cdots + (-1)^n\,t^n \\ &= (-1)^n \cdot (t-\lambda_1)^{m_1}\,(t-\lambda_2)^{m_2}\cdots(t-\lambda_k)^{m_k} \end{aligned}$$
>
>   has linear factors with each eigenvalue $\lambda_k \in F$. Then there exists an orthonormal basis $\beta = \{v_1,\,v_2,\,\ldots,\,v_n\}$ for V such that the matrix $\big[\mathsf{T}\big]_\beta^\beta$ is upper triangular.
> - Assume that V is a vector space over $F = \mathbb{C}$. Then T is normal if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.

Let V be a finite dimensional vector space over either $F = \mathbb{R}$ or $F = \mathbb{C}$, and fix an inner product $\langle\cdot,\,\cdot\rangle : \mathsf{V}\times\mathsf{V} \to F$. A linear operator $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ is said to be *self-adjoint* or *Hermitian* if $\mathsf{T} = \mathsf{T}^*$.

In practice, we choose $F = \mathbb{R}$, so that a self-adjoint operator T corresponds to a symmetric matrix $A = \left[\mathsf{T}\right]_\beta^\beta$, that is, $A = A^t$.

Theorem 6.17. Let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be a linear operator on a n-dimensional inner product space V.

- T is normal.
- Every eigenvalue λ of T is a real number, that is, $\lambda \in \mathbb{R}$.
- The characteristic polynomial of T splits, that is, the polynomial
$$f(t) = \det\left(\left[\mathsf{T}\right]_\beta^\beta - t\, I_n\right) = \det(\mathsf{T}) + \cdots + (-1)^n\, t^n$$
$$= (-1)^n \cdot (t - \lambda_1)^{m_1}\, (t - \lambda_2)^{m_2} \cdots (t - \lambda_k)^{m_k}$$
has linear factors.
- Assume that V is a vector space over $F = \mathbb{R}$. Then T is self-adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.