Lecture 1, Jan 11, 2022

Review: Matrices

- $A = [a_{ij}] \in {}^m\mathbb{R}^n$ is a general $m \times n$ matrix in the reals
  - ${}^m\mathbb{R}$ is the set of all $m \times 1$ real matrices (i.e. columns)
  - $\mathbb{R}^n$ is the set of all $1 \times n$ real matrices (i.e. rows)
  - We can also have matrices of $\mathbb{C}, \mathbb{Z}, \mathbb{Q}$, etc
- Matrix multiplication: $A \in {}^m\mathbb{R}^n, B \in {}^n\mathbb{R}^p \implies C = AB = [c_{ij}] \in {}^m\mathbb{R}^p, c_{ij} = \displaystyle\sum_{k=1}^{n} a_{ik}b_{kj}$
- Transpose swaps rows and columns: $A \in {}^m\mathbb{R}^n \implies A^T \in {}^n\mathbb{R}^m$
- Trace is the sum of the main diagonal $A \in {}^n\mathbb{R}^n \implies \operatorname{tr} A = \displaystyle\sum_{i=1}^{n} a_{ii}$ (for square matrices only)
- Determinant $\det A$ for square matrices
- Inverse A^{-1} also for square matrices; exists only when $\det A \neq 0$ (i.e. matrix is full rank); $AA^{-1} = A^{-1}A = I$
  - Pseudoinverses exist for nonsquare matrices
  - For square matrices $AB = I \implies BA = I$
- $(A^T)^{-1} = (A^{-1})^T$ so sometimes we denote $(A^{-1})^T = A^{-T}$
- Symmetric matrices $A^T = A$, skew symmetric (or anti-symmetric) $A^T = -A$
- Identity denoted as a boldface 1
- Matrix addition is associative and commutative; matrix multiplication is associative and not commutative
- We have closure under matrix addition and scalar multiplication; that is after adding two matrices and multiplying by a scalar we still get a matrix in the same dimensions

Lecture 2, Jan 14, 2022

Vector Spaces

- A *vector space* is the generalized concept of a vector that satisfies the usual rules of vector arithmetic
- Fundamental abstract operations addition $+$ and scalar multiplication $\cdot$ can be defined in any way, not just the common component-wise way
  - If it can be defined in any way, what makes a definition meaningful? When does it make sense?
- Definition: A *vector space* $\mathcal{V}$ over a field Γ of elements $\{\alpha, \beta, \gamma, \cdots\}$ called *scalars*, is a set of elements $\{u, v, w, \cdots\}$ such that the following *axioms* are satisfied:
  1. Vector addition denoted $u + v$ satisfies, for all $u, v, w \in \mathcal{V}$ (properties AI - AIV):
     1. Closure: $u + v \in \mathcal{V}$
     2. Associativity: $(u + v) + w = u + (v + w)$
     3. Existence of zero or null vector $0 \in \mathcal{V}$ such that $u + 0 = u$
     4. Existence of a negative or additive inverse $-u \in \mathcal{V}$ such that $u + (-u) = 0$
  2. Scalar multiplication denoted αu, such that for all $u, v \in \mathcal{V}$ and $\alpha, \beta \in \Gamma$ (properties MI - MIV):
     1. Closure: $\alpha u \in \mathcal{V}$
     2. Associativity: $\alpha(\beta u) = (\alpha\beta)u$
     3. Distributivity: $(\alpha + \beta)u = \alpha u + \beta u$, and $\alpha(u + v) = \alpha u + \alpha v$
     4. Unitary: For the identity $1 \in \Gamma$, $1u = u$
- Note that these properties imply commutativity for vector addition (will prove in a following lecture)
- A field Γ is a commutative group that has two operations, addition and multiplication (between scalars), and has a set of elements such that:
  1. Γ is commutative under addition
  2. Γ is commutative under multiplication excluding zero
  3. Multiplication is distributive over addition
- For us the field is almost always going to be $\mathbb{R}$; other examples of fields include the rationals, the complex numbers, etc

- A group is a set of elements $\{x, y, z, \cdots\}$ and a binary operation xy such that the operation is closed, associative, and there exists an inverse and identity for this operation; commutative groups additionally have $xy = yx$
- Matrices are an example of a vector space since they satisfy all of the requirements, so we can think of matrices as vectors
- Formally we would say $\mathcal{V}$ is a vector space over the field Γ under vector addition $+$ and scalar multiplication $\cdot$; as a shorthand we just say $\mathcal{V}$ is a vector space

Lecture 3, Jan 17, 2022

Vector Spaces, Continued

- Consider: $\{\ (x, y) \mid x, yin\mathbb{R}\ \}$ with the operations defined as $(x_1, y_1) + (x_2, y_2) \equiv (x_1 + x_2, y_1 + y_2 + 1)$ and $\alpha(x, y) \equiv (\alpha x, \alpha y + \alpha - 1)$
  - Zero is $(0, -1)$
  - Inverse is $(-x, -y - 2)$
  - Actually distributive
  - Since all axioms hold this is actually a vector space

Lecture 4, Jan 18, 2022

Commutativity and Other Properties of Vector Spaces

- What about the associative property $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$?
  - This can be proven from the other properties, but first we will start with some other axioms
- There are some axioms that are one sided such as $\mathcal{AIII}$ and $\mathcal{AIV}$ (additive identity and inverse); we will prove that these are two sided under the other axioms
- Proposition I: For every $\boldsymbol{u}, -\boldsymbol{u} \in \mathcal{V}, -\boldsymbol{u} + \boldsymbol{u} = \boldsymbol{0}$ (i.e. property $\mathcal{AIV}$ but commutative)
  - Proof: $$\begin{aligned} -\boldsymbol{u} + \boldsymbol{u} &= (-\boldsymbol{u} + \boldsymbol{u}) + \boldsymbol{0} && \mathcal{AIII} \\ &= (-\boldsymbol{u} + \boldsymbol{u}) + (-\boldsymbol{u} + (-(-\boldsymbol{u}))) && \mathcal{AIV} \\ &= -\boldsymbol{u} + (\boldsymbol{u} + (-\boldsymbol{u})) + (-(-\boldsymbol{u})) && \mathcal{AII} \\ &= -\boldsymbol{u} + \boldsymbol{0} + (-(-\boldsymbol{u})) && \mathcal{AIV} \\ &= -\boldsymbol{u} + (-(-\boldsymbol{u})) && \mathcal{AIII} \\ &= \boldsymbol{0} && \mathcal{AIV} \end{aligned}$$
  - Thus $\mathcal{AIV}$ is commutative, and we can say that the additive inverse of $-\boldsymbol{u}$ is just $\boldsymbol{u}$
- Proposition II: For every $u \in \mathcal{V}, \boldsymbol{0} + \boldsymbol{u} = \boldsymbol{u}$ (i.e. property $\mathcal{AIII}$ but commutative)
  - Proof: $$\begin{aligned} \boldsymbol{0} + \boldsymbol{u} &= (\boldsymbol{u} + (-\boldsymbol{u})) + \boldsymbol{u} && \mathcal{AIV} \\ &= \boldsymbol{u} + (-\boldsymbol{u} + \boldsymbol{u}) && \mathcal{AII} \\ &= \boldsymbol{u} + \boldsymbol{0} && \text{Prop. I} \\ &= \boldsymbol{u} && \mathcal{AIII} \end{aligned}$$
- Theorem I: Cancellation theorem: If $\boldsymbol{u} + \boldsymbol{w} = \boldsymbol{v} + \boldsymbol{w}$ then $\boldsymbol{u} + \boldsymbol{v}$ for any $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$ (this also applies for $\boldsymbol{w} + \boldsymbol{u} = \boldsymbol{w} + \boldsymbol{v}$)
  - Proof: $$\begin{aligned} \boldsymbol{u} &= \boldsymbol{u} + \boldsymbol{0} && \mathcal{AIII} \\ &= \boldsymbol{u} + (\boldsymbol{w} + -\boldsymbol{w}) && \mathcal{AIV} \\ &= (\boldsymbol{u} + \boldsymbol{w}) + (-\boldsymbol{w}) && \mathcal{AII} \\ &= (\boldsymbol{v} + \boldsymbol{w}) + (-\boldsymbol{w}) && \text{given} \\ &= \boldsymbol{v} + (\boldsymbol{w} + (-\boldsymbol{w})) && \mathcal{AII} \\ &= \boldsymbol{v} + \boldsymbol{0} && \mathcal{AIV} \\ &= \boldsymbol{v} && \mathcal{AIII} \end{aligned}$$

- Note: a *theorem* and *proposition* are basically the same thing here, but typically theorem is used for bigger results
- Define subtraction: $\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v})$
- Proposition III:
  1. The zero $\mathbf{0} \in \mathcal{V}$ is unique
     - Proof: Let $\mathbf{0}'$ be another zero, then $\boldsymbol{u} + \mathbf{0}' = \boldsymbol{u} = \boldsymbol{u} + \mathbf{0}$ so by the cancellation theorem, $\mathbf{0}' = \mathbf{0}$
  2. The inverse is unique
  3. $-(-\boldsymbol{u}) = \boldsymbol{u}$
- Proposition IV: For $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}, \boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
  - Proof:
$$\begin{aligned}
\boldsymbol{u} + \boldsymbol{v} &= \mathbf{0} + (\boldsymbol{u} + \boldsymbol{v}) + \mathbf{0} && \text{Prop. II and } \mathcal{AIII} \\
&= (-\boldsymbol{v} + \boldsymbol{v}) + (\boldsymbol{u} + \boldsymbol{v}) + (\boldsymbol{u} + (-\boldsymbol{u})) && \text{Prop. I and } \mathcal{AIV} \\
&= -\boldsymbol{v} + ((\boldsymbol{v} + \boldsymbol{u}) + (\boldsymbol{v} + \boldsymbol{u})) + (-\boldsymbol{u}) && \mathcal{AII} \\
&= -\boldsymbol{v} + (1(\boldsymbol{v} + \boldsymbol{u}) + 1(\boldsymbol{v} + \boldsymbol{u})) + (-\boldsymbol{u}) && \mathcal{MIV} \\
&= -\boldsymbol{v} + (1 + 1)(\boldsymbol{v} + \boldsymbol{u}) + (-\boldsymbol{u}) && \mathcal{MIII} \\
&= -\boldsymbol{v} + ((1 + 1)\boldsymbol{v} + (1 + 1)\boldsymbol{u}) + (-\boldsymbol{u}) && \mathcal{MIII} \\
&= -\boldsymbol{v} + (1\boldsymbol{v} + 1\boldsymbol{v} + 1\boldsymbol{u} + 1\boldsymbol{u}) + (-\boldsymbol{u}) && \mathcal{MIII} \\
&= -\boldsymbol{v} + (\boldsymbol{v} + \boldsymbol{v} + \boldsymbol{u} + \boldsymbol{u}) + (-\boldsymbol{u}) && \mathcal{MIV} \\
&= (-\boldsymbol{v} + \boldsymbol{v}) + \boldsymbol{v} + \boldsymbol{u} + (\boldsymbol{u} + (-\boldsymbol{u})) && \mathcal{AII} \\
&= \mathbf{0} + \boldsymbol{v} + \boldsymbol{u} + \mathbf{0} && \text{Prop. I and } \mathcal{AIV} \\
&= \boldsymbol{v} + \boldsymbol{u} && \text{Prop. II and } \mathcal{AIII}
\end{aligned}$$

Lecture 5, Jan 24, 2022

More Vector Space Properties

- Proposition V: Properties of zero: For all $\boldsymbol{v} \in \mathcal{V}$ and all $\alpha \in \Gamma$:
  1. $0\boldsymbol{v} = \mathbf{0}$
     - $0\boldsymbol{v} = 0\boldsymbol{v} + \mathbf{0}$ by $\mathcal{AIII}$
     - $0\boldsymbol{v} = (0 + 0)\boldsymbol{v} = 0\boldsymbol{v} + 0\boldsymbol{v}$ by $\mathcal{MIII}(a)$ and scalar addition properties
     - By the transitive property $0\boldsymbol{v} + \mathbf{0} = 0\boldsymbol{v} + 0\boldsymbol{v}$, then by the cancellation theorem $0\boldsymbol{v} = \mathbf{0}$
  2. $\alpha\mathbf{0} = \mathbf{0}$
     - $\alpha\mathbf{0} = \alpha(\mathbf{0} + \mathbf{0}) = \alpha\mathbf{0} + \alpha\mathbf{0}$
     - Rest of the proof follows like above
  3. If $\alpha\boldsymbol{v} = \mathbf{0}$ then either $\alpha = 0$ or $\boldsymbol{v} = \mathbf{0}$
     - Either $\alpha = 0$ or $\alpha \neq 0$; if $\alpha = 0$ then $0\boldsymbol{v} = \mathbf{0}$ follows by 1, so we only need to consider $\alpha \neq 0$
     - $$\begin{aligned}
\boldsymbol{v} &= 1\boldsymbol{v} && \mathcal{MIV} \\
&= (\alpha^{-1}\alpha)\boldsymbol{v} && \text{Properties of scalars} \\
&= \alpha^{-1}(\alpha\boldsymbol{v}) && \mathcal{MII} \\
&= \alpha^{-1}\mathbf{0} && \text{Given} \\
&= \mathbf{0} && \text{Prop. V.2}
\end{aligned}$$
     - Therefore either $\alpha = 0$, or if $\alpha \neq 0$, then $\boldsymbol{v} = \mathbf{0}$
- Proposition VI: For all $\boldsymbol{v} \in \mathcal{V}$ and $\alpha \in \Gamma$, $(-\alpha)\boldsymbol{v} = -(\alpha\boldsymbol{v}) = \alpha(-\boldsymbol{v})$
  - $$\begin{aligned}
\alpha\boldsymbol{v} + (-\alpha\boldsymbol{v}) &= (\alpha - \alpha)\boldsymbol{v} && \mathcal{MIII}(a) \\
&= 0\boldsymbol{v} && \text{Properties of scalars} \\
&= \mathbf{0} && \text{Prop. V.1}
\end{aligned}$$
  - Since $\alpha\boldsymbol{v} + (-(\alpha\boldsymbol{v})) = \mathbf{0}$ by $\mathcal{AIII}$, by the transitive property and cancellation $-(\alpha\boldsymbol{v}) = (-\alpha)\boldsymbol{v}$
  - $$\begin{aligned}
\alpha\boldsymbol{v} + \alpha(-\boldsymbol{v}) &= \alpha(\boldsymbol{v} - \boldsymbol{v}) && \mathcal{MIII}(b) \\
&= \alpha\mathbf{0} && \mathcal{AIV} \\
&= \mathbf{0} && \text{Prop. V.2}
\end{aligned}$$

- It follows then that $\alpha(-\boldsymbol{v}) = -(\alpha\boldsymbol{v}) = (-\alpha\boldsymbol{v})$
- Consider $\alpha = 1$, then $-(1\boldsymbol{v}) = -\boldsymbol{v} = (-1)\boldsymbol{v}$, so the additive inverse is always -1 times the vector!

Lecture 6, Jan 25, 2022

Subspaces

- A subset $\mathcal{U}$ of $\mathcal{V}$ is a subspace of $\mathcal{V}$ iff $\mathcal{U}$ is itself a vector space over the same field Γ with the same vector addition and scalar multiplication operations of $\mathcal{V}$
  - $X \subseteq Y \iff \forall x \in X \implies x \in Y$
  - In this case the subset is not strict, i.e. $\mathcal{U} = \mathcal{V}$ is allowed
  - Every $\mathcal{V}$ has two subspaces, the space itself, and the subspace of only zero: $\mathcal{U} = V$ and $\mathcal{U} = \{ \mathbf{0} \}$
  - Sometimes the notation $\mathcal{U} \sqsubseteq \mathcal{V}$ is used
- Theorem 1: Subspace test: $\mathcal{U} \sqsubseteq \mathcal{V}$ iff for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{U}$ and all $\alpha \in \Gamma$:
  1. Zero: $\exists \mathbf{0} \in \mathcal{U} \ni \boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$
  2. Closure under addition: $\boldsymbol{u} + \boldsymbol{v} \in \mathcal{U}$
  3. Closure under scalar multiplication: $\alpha\boldsymbol{u} \in \mathcal{U}$
- Proof of the subspace test:
  - $\mathcal{U} \sqsubseteq \mathcal{V} \implies (\mathcal{SI}, \mathcal{SII}, \mathcal{SIII})$: By definition $\mathcal{U}$ is a vector space, therefore it automatically satisfies all 3 axioms
  - $(\mathcal{SI}, \mathcal{SII}, \mathcal{SIII}) \implies \mathcal{U} \sqsubseteq \mathcal{V}$:
    * $\mathcal{AI}$: Implied by $\mathcal{SII}$
    * $\mathcal{AII}$: Automatically true since $\boldsymbol{u} \in \mathcal{U} \implies \boldsymbol{u} \in \mathcal{V}$ and the addition operator is associative in $\mathcal{V}$ (i.e. inherited from $\mathcal{V}$)
    * $\mathcal{AIII}$: Implied by $\mathcal{SI}$
    * $\mathcal{AIV}$: We have proven previously that $(-1)\boldsymbol{u}$ is the additive inverse of $\boldsymbol{u}$; we also know $(-1)\boldsymbol{u} \in \mathcal{U}$ by $\mathcal{SIII}$, so an inverse exists
    * $\mathcal{MI}$: Implied by $\mathcal{SIII}$
    * $\mathcal{MII}$ – $\mathcal{MIII}$: Inherited from $\mathcal{V}$
    * $\mathcal{MIV}$: $1\boldsymbol{u} \in \mathcal{U}$ by $\mathcal{SIII}$ and $\boldsymbol{u} \in \mathcal{V}$ so $1\boldsymbol{u} = \boldsymbol{u} \in \mathcal{U}$
- Example: $\operatorname{im} A = \{ \boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \forall \boldsymbol{x} \in {}^n\mathbb{R} \} \subseteq {}^m\mathbb{R}$ for $\boldsymbol{A} \in {}^m\mathbb{R}^n$
  - Since ${}^m\mathbb{R}$ is a vector space over $\mathbb{R}$ we only need to do the subspace test
  - $\mathcal{SI}$: Satisfied since $\mathbf{0} = \boldsymbol{A}\mathbf{0} \implies \mathbf{0} \in \operatorname{im} \boldsymbol{A}$
  - $\mathcal{SII}$: $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \operatorname{im} \boldsymbol{A} \implies \boldsymbol{y}_1 = \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{y}_2 = \boldsymbol{A}\boldsymbol{x}_2 \implies \boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2)$
  - $\mathcal{SIII}$: $\boldsymbol{y} \in \operatorname{im} \boldsymbol{A} \implies \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \implies \alpha\boldsymbol{y} = \alpha(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}(\alpha\boldsymbol{x}) \in \operatorname{im} \boldsymbol{A}$
- Example: $\ker \boldsymbol{A} = \{ \boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \mathbf{0} \}$ for $\boldsymbol{A} \in {}^m\mathbb{R}^n$ (kernel or null space of A)
  - $x \in {}^n\mathbb{R}$ so we can apply the subspace test
  - $\mathcal{SI}$: $\mathbf{0} \in \ker \boldsymbol{A}$ because $\boldsymbol{A}\mathbf{0} = \mathbf{0}$
  - $\mathcal{SI}$: $\boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0} \implies \boldsymbol{x}_1 + \boldsymbol{x}_2 \in \ker \boldsymbol{A}$
  - $\mathcal{SIII}$: $\boldsymbol{A}(\alpha\boldsymbol{x}) = \alpha(\boldsymbol{A}\boldsymbol{x}) = \alpha\mathbf{0} = \mathbf{0} \implies \alpha\boldsymbol{x} \in \ker \boldsymbol{A}$

Lecture 7, Jan 28, 2022

Linear Combination and Span

- Definition: A vector $\boldsymbol{v} \in \mathcal{V}$ is a linear combination of $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$ if and only if it can be written as $\boldsymbol{v} = \displaystyle\sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j$ for $\lambda_j \in \Gamma$
  - Note the use of $\subset$ instead of $\subseteq$ because for now we want to keep the set finite
- Definition: The *span* of $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$ is denoted: $\operatorname{span} \{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} = \left\{ \boldsymbol{v} \;\middle|\; \boldsymbol{v} = \displaystyle\sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j, \forall \lambda_j \in \Gamma \right\}$, i.e. all the vectors that can be written as a linear combination of this set of vectors

- Example: ${}^3\mathbb{R} = \text{span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}$
  - $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ is the *spanning set* of vectors (for now, this set will be finite, but the span itself is infinite)
- Proposition I: The span of $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \in \mathcal{V}$ is a subspace of $\mathcal{V}$
  - Proof:
    * $\mathcal{SI}$: $\mathbf{0} = \sum_{j=1}^{n} 0\boldsymbol{v}_j$ therefore $\mathbf{0}$ in this subset
    * $\mathcal{SII}$: Let $\boldsymbol{u} \in \mathcal{V} = \sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j$ and $\boldsymbol{w} \in \mathcal{V} = \sum_{j=1}^{n} \beta_j \boldsymbol{v}_j$ then $\boldsymbol{u} + \boldsymbol{w} = \sum_{j=1}^{n} (\alpha_j + \beta_j)\boldsymbol{v}_j$
    * $\mathcal{SIII}$: $\boldsymbol{u} \in \mathcal{V} = \sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j$, then $\lambda \boldsymbol{u} = \lambda \sum_{j=1}^{n} \alpha_j \boldsymbol{v}_j = \sum_{j=1}^{n} (\lambda \alpha_j)\boldsymbol{v}_j$

Lecture 8, Jan 31, 2022

Equivalence of Spans

- We can show equivalence of sets $U = V$ by showing $U \subseteq V$ and $V \subseteq U$
- To show equivalence of spans, we do the same and show both spans are subsets of the other
  - It is sufficient to show that each member of the spanning set is in the other span; if $\boldsymbol{u}_i = \sum_{k=1}^{n} \boldsymbol{\alpha}_{ik} v_k$, then $\boldsymbol{u} = \sum_{i=1}^{m} \lambda_i \boldsymbol{u}_i = \sum_{i=1}^{m} \lambda_i \left(\sum_{k=1}^{n} \alpha_{ik} \boldsymbol{v}_k \right) = \sum_{k=1}^{n} \left(\sum_{i=1}^{m} \right) \boldsymbol{v}_k = \sum_{k=1}^{n} \mu_k \boldsymbol{v}_k \implies \boldsymbol{u} \in \text{span}\{ \boldsymbol{v}_j \}$
- Example: $\text{span}\{ \boldsymbol{u}, \boldsymbol{v} \} = \text{span}\{ \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \}$?
  - $\{ \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \} \subseteq \text{span}\{ \boldsymbol{u}, \boldsymbol{v} \}$ since they're both linear combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$
  - $\{ \boldsymbol{u}, \boldsymbol{v} \} \subseteq \text{span}\{ \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \}$ since $\boldsymbol{u} = \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) + \frac{1}{2}(\boldsymbol{u} - \boldsymbol{v})$ and $\boldsymbol{v} = \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) - \frac{1}{2}(\boldsymbol{u} - \boldsymbol{v})$
- Proposition II: Let $\mathcal{U} = \text{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \sqsubseteq \mathcal{V}$. If $\mathcal{W}$ is a subspace of $\mathcal{V}$ containing the vectors $\text{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ then $\mathcal{U} \sqsubseteq \mathcal{W}$.
  - Any vector in $\mathcal{U}$ is a linear combination of those vectors, and since those vectors are in $\mathcal{W}$, $\mathcal{W}$ contains all linear combinations of those vectors

Linear Independence

- Linear independence: A set of vectors $\text{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ is *linearly independent* if and only if $\sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j = \mathbf{0} \implies \lambda_j = 0$
  - Example: $\left\{ \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix} \right\}$ are independent: $\lambda_1 \begin{bmatrix}1\\0\end{bmatrix} + \lambda_2 \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \implies \begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \implies \lambda_1 = \lambda_2 = 0$
  - Example: $\left\{ \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix} \right\}$ are not independent: $\lambda_1 \begin{bmatrix}1\\0\end{bmatrix} + \lambda_2 \begin{bmatrix}0\\1\end{bmatrix} + \lambda_3 \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$ can be satisfied with $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 1 \\ \lambda_3 = -1 \end{cases}$
- Proposition I: If $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$ is linearly independent and $\boldsymbol{v} = \sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j$ for all $\boldsymbol{v} \in \mathcal{V}$ then λ_j are uniquely determined, i.e. there is only one way to construct any vector

- Proof: Assume that λ_j are not uniquely determined. Let $\boldsymbol{v} = \sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j = \sum_{j=1}^{n} \mu_j \boldsymbol{v}_j$ then $\mathbf{0} = \boldsymbol{v} - \boldsymbol{v} = \sum_{j=1}^{n} (\lambda_j - \mu_j) = \boldsymbol{v}_j$ and because the set is linearly independent $\lambda_j - \mu_j = 0 \implies \lambda_j = \mu_j$, so λ_j are uniquely determined
- This generalizes to any kind of vector, e.g. functions
  - e.g. to show $\{ \sin x, \cos x \}$ are linearly independent we show $\lambda_1 f + \lambda_2 g = z \implies \lambda_1 = \lambda_2 = 0$ where $z : \mathbb{R} \mapsto \{ 0 \}$
    * We want to show $\lambda_1 \cos x + \lambda_2 \sin x = 0 \forall x \in \mathbb{R}$; we can consider $x = 0 \implies \lambda 1 = 0$, and $x = \frac{\pi}{2} \implies \lambda_2 = 0$

Lecture 9, Feb 1, 2022

Basis

- Fundamental Theorem of Linear Algebra: Let $\mathcal{V}$ be a vector space spanned by n vectors. If a set of m vectors from $\mathcal{V}$ is linearly independent, then $m \leq n$.
  - This is equivalent to saying if $m > n$ then any set of m vectors from $\mathcal{V}$ is linearly dependent (this is the contrapositive statement: if $A \implies B$, then $\neg B \implies \neg A$)
  - Proof by contraposition: Let $m > n$, we show that this implies a set of m vectors is dependent.
    * Consider a set of m vectors $\{ \boldsymbol{u}_1, \cdots, \boldsymbol{u}_m \}$ and let $\operatorname{span} \{ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \} = \mathcal{V}$
    * Since $\boldsymbol{u}_j \in \mathcal{V}$, $\boldsymbol{u}_j = \sum_{i=1}^{n} a_{ij} \boldsymbol{v}_i$ and so
    $$\sum_{j=1}^{m} x_j \boldsymbol{u}_j = \sum_{j=1}^{m} x_j \left(\sum_{i=1}^{n} a_{ij} \boldsymbol{v}_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} x_j \right) \boldsymbol{v}_i$$
    * Set $\sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} x_j \right) \boldsymbol{v}_i = \mathbf{0}$; this will be satisfied if each $\sum_{j=1}^{m} a_{ij} x_j = 0$; this is a set of linear equations $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ where $\boldsymbol{A}$ is $n \times m$; since we have $m > n$ there are infinite number of solutions to this system, i.e. there exist a nontrivial solution, therefore not all x_j have to be 0, so the set $\{ \boldsymbol{u}_1, \cdots, \boldsymbol{u}_m \}$ is linearly dependent
- Define the *basis* for a vectors space $\mathcal{V}$ to be a set of vectors $\{ \boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n \}$ that are linearly independent and $\operatorname{span} \{ \boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m \} = \mathcal{V}$
  - Every basis for a given vector space contains the same number of vectors:
    * Let $E = \{ \boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n \}$ and $F = \{ \boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n \}$ be bases
    * Consider E to be linearly independent and F to span $\mathcal{V}$, then by the fundamental theorem $n \leq m$; consider F to be linearly independent and E to span $\mathcal{V}$, then by the fundamental theorem $m \leq n$, therefore $m = n$
  - We say that a basis *generates* $\mathcal{V}$
- Definition: The *dimension* of a vector space $\mathcal{V}$, denoted $\dim \mathcal{V}$, is the number of vectors in any of its bases
  - Note: Define $\dim \{ \mathbf{0} \} = 0$

Lecture 10, Feb 4, 2022

Basis Continued

- In general $\dim {}^n\mathbb{R} = \dim \mathbb{R}^n = n$ and $\dim {}^m\mathbb{R}^n = mn$

- The standard basis is the set $\left\{ \begin{bmatrix}1\\0\\0\\\vdots\end{bmatrix}, \begin{bmatrix}0\\1\\0\\\vdots\end{bmatrix}, \begin{bmatrix}0\\0\\1\\\vdots\end{bmatrix}, \cdots \right\}$
- Example: Consider the space of skew-symmetric matrices $\mathcal{U} = \{ \boldsymbol{S} \mid \boldsymbol{S} = -\boldsymbol{S}^T, \boldsymbol{S} \in {}^3\mathbb{R}^3 \}$
  - $\boldsymbol{S} = \boldsymbol{S}^T$ means the diagonal is forced to be zero
  - We can now write $\boldsymbol{S} = \begin{bmatrix}0 & a & b\\ -a & 0 & c\\ -b & -c & 0\end{bmatrix} = a\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix} + b\begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ -1 & 0 & 0\end{bmatrix} + c\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1 & 0\end{bmatrix}$
  - Since those 3 matrices span $\mathcal{U}$ and are linearly independent they form a basis, therefore $\dim \mathcal{U} = 3$
- Let $\mathcal{V}$ be a finite-dimensional vector space with $\dim \mathcal{V} = n$, then
  1. A linearly independent set of vectors in $\mathcal{V}$ can at most contain n vectors
  2. A spanning set for $\mathcal{V}$ must contain at least n vectors
- We can add vectors to any linearly independent set until we have $\mathcal{V} = n$ vectors; we can take away vectors from any spanning set until we have n vectors; at n vectors, we can have a spanning set that is linearly independent
  - Sometimes referred to as the rule of the extreme middle
- A basis characterizes a vector space

Lecture 11, Feb 7, 2022

Minimal Spanning Sets

- Theorem I: Let $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$. For every v_k (where $k = 1, 2, \cdots n$), $\operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \cdots \boldsymbol{v}_n \} \subsetneq \operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ iff $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ is linearly independent
  - i.e. if we have a set of linearly independent vectors and we take a vector out, the resulting span is always a strict subset (gets smaller)
  - Corollary (contrapositive): Let $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$, then for at least one v_k, $\operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \cdots \boldsymbol{v}_n \}$ $\operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ if and only if $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ is linearly dependent
    - Note the original theorem is for every v_k but the contrapositive is for at least one v_k
  - Proof: At least one $\boldsymbol{v}_k$ implies linear dependence:
    - Let there be one $\boldsymbol{v}_k$ such that $\operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \cdots \boldsymbol{v}_n \} = \operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$
    - Then $\boldsymbol{v}_k \in \operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \cdots \boldsymbol{v}_n \}$
    - $\boldsymbol{v}_k = \sum_{\substack{i=1\\(i \neq k)}}^{n} \lambda_i \boldsymbol{v}_i \implies \lambda_1 \boldsymbol{v}_1 + \cdots + (-1)\boldsymbol{v}_k + \cdots + \lambda_n \boldsymbol{v}_n = \boldsymbol{0}$
    - Therefore the set $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \}$ is not linearly independent because the coefficient on $\boldsymbol{v}_k$ can be nonzero
  - Proof: Linear dependence implies existence of $\boldsymbol{v}_k$:
    - $\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i = \boldsymbol{0}$ has at least one $\lambda_k \neq 0$; take the vector associated with this λ to be $\boldsymbol{v}_k$
    - $\lambda_k \boldsymbol{v}_k = -\lambda_1 \boldsymbol{v}_1 - \cdots - \lambda_{k-1} \boldsymbol{v}_{k-1} - \lambda_{k+1} \boldsymbol{v}_{k+1} - \cdots - \lambda_n \boldsymbol{v}_n$
    - $\boldsymbol{v}_k = -\frac{\lambda_1}{\lambda_{k\,1}} \boldsymbol{v}_1 - \cdots - \frac{\lambda_{k-1}}{\lambda_k} \boldsymbol{v}_{k-1} - \frac{\lambda_{k+1}}{\lambda_k} \boldsymbol{v}_{k+1} - \cdots - \frac{\lambda_n}{\lambda_k} \boldsymbol{v}_n$
    - Therefore $\boldsymbol{v}_k \in \operatorname{span}\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_{k+1}, \cdots \boldsymbol{v}_n \}$ so the span with $\boldsymbol{v}_k$ is the same as the span without $\boldsymbol{v}_k$
- Any minimum spanning set of a vector space is also a maximum independent set of that vector space
  - Taking any vector out of a set of linearly independent vectors loses information
- Theorem IV: Let $\{ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \} \subset \mathcal{V}$ be linearly independent; then for another $\boldsymbol{v} \in \mathcal{V}$, $\{ \boldsymbol{v}, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \}$ is linearly independent if and only if $\boldsymbol{v} \notin \operatorname{span}\{ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \}$; i.e. we can add a vector to a linearly independent set and keep it linearly independent if this vector is not already in the span
  - Contrapositive: If $\boldsymbol{v} \in \operatorname{span}\{ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \}$ then $\{ \boldsymbol{v}, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \}$ is linearly dependent
  - Proof:

  * $\boldsymbol{v} \in \operatorname{span}\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\} \implies \boldsymbol{v} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \implies (-1)\boldsymbol{v} + \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n = \boldsymbol{0}$ therefore the set is linearly dependent
  * $\{\boldsymbol{v}, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ linearly dependent means $\lambda\boldsymbol{v} + \lambda_1 \boldsymbol{v}_1 + \cdots + \lambda_n \boldsymbol{v}_n = \boldsymbol{0}$ and not all λs equal zero
    - First we need to show $\lambda \neq 0$: If $\lambda = 0$, that means the rest of the λ_i have to be 0, which would mean $\{\boldsymbol{v}, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ is linearly independent, creating a contradiction
    - Since $\lambda \neq 0$, $\boldsymbol{v} = -\frac{\lambda_1}{\lambda}\boldsymbol{v}_1 - \cdots - \frac{\lambda_n}{\lambda}\boldsymbol{v}_n$

Lecture 12, Feb 8, 2022

Existence of Bases

- Theorem V: Let $\mathcal{V}$ be spanned by a finite set of vectors; then every linearly independent set in $\mathcal{V}$ can be extended to a basis for $\mathcal{V}$ (note we assume the set is not the zero set)
  - Proof by construction:
    1. Start with a linearly independent set $S_k = \{\boldsymbol{v}_1, \cdots \boldsymbol{v}_k\} \subset \mathcal{V}$
    2. $\operatorname{span} S_k = \mathcal{V}$ or $\operatorname{span} S_k \neq \mathcal{V}$; if $\operatorname{span} S_k = \mathcal{V}$ then we have a linearly independent spanning set, which is a basis, so we're done
    3. Otherwise, $\exists(\boldsymbol{v}_{k+1} \in \mathcal{V}) \notin \operatorname{span} S_k$; by Theorem IV, $S_{k+1} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}\}$ is linearly independent
    4. If $\operatorname{span} S_k = \mathcal{V}$ then we have a basis; otherwise, repeat the previous step until we eventually get a basis
       * Because $\mathcal{V}$ is spanned by a finite set of vectors, this will always result in a basis
       * If it doesn't result in a basis then we'll end up with a set of linearly independent vectors that don't span $\mathcal{V}$ but has more vectors than the finite set that spans $\mathcal{V}$, which violates the fundamental theorem
- Theorem V gives a *maximally linearly independent set*
- Theorem VII: Let $\mathcal{V}$ be spanned by a finite set of vectors; then any spanning set for $\mathcal{V}$ can be reduced to a basis (i.e. it contains a basis)
  - Proof by construction:
    1. Start with a spanning set $\operatorname{span} S_p = \mathcal{V}$ where $S_p = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\} \subset \mathcal{V}$
    2. S_p is either linearly independent or not; if it is then S_p is a basis and we're done
    3. Otherwise, by Theorem I Corollary, $\exists \boldsymbol{v}_p \in S_p$ such that $\operatorname{span} S_{p-1} = \mathcal{V}$ where $S_{p-1} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{p-1}\}$ (renumber the vectors such that v_p is that vector)
    4. If S_{p-1} is linearly independent then we have a basis; otherwise repeat the previous step until we eventually get a basis
       * This process must stop because eventually we get a set with just 1 vector which will be linearly independent
- Theorem VII gives a *minimally spanning set*
- Bases can be thought as minimally spanning sets or maximally independent sets
- Theorem VIII: Let $\mathcal{V}$ be such that $\dim \mathcal{V} = n$; then:
  1. Any linearly independent set of n vectors is a basis
  2. Any spanning set of n vectors is a basis
- Theorem VI: Let $\mathcal{U}, \mathcal{W} \sqsubseteq \mathcal{V}$, then:
  1. $\mathcal{U}, \mathcal{W}$ are finite dimensional with dimensions less than or equal to $\mathcal{V}$
  2. If $\mathcal{U} \subseteq \mathcal{W}$ then $\dim \mathcal{U} \leq \dim \mathcal{W}$
  3. $\mathcal{U} \subseteq \mathcal{W} \land \dim \mathcal{U} = \dim \mathcal{W} \implies \mathcal{U} = \mathcal{W}$

Lecture 13, Feb 11, 2022

Null, Column, and Row Space

- The null space is defined as null $\boldsymbol{A} = \{ \boldsymbol{x} \in {}^n\mathbb{R} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \} \subseteq \mathbb{R}$
- Consider $\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in {}^m\mathbb{R}^n$; $\boldsymbol{A}$ can be expressed as a set of rows $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix}, \boldsymbol{r}_k \in \mathbb{R}^n$ or a set of columns $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_n \end{bmatrix}, \boldsymbol{c}_j \in {}^m\mathbb{R}$
- Define the *row space* of $\boldsymbol{A}$ as $\operatorname{row} \boldsymbol{A} = \operatorname{span}\{ \boldsymbol{r}_1, \cdots \boldsymbol{r}_m \} \sqsubseteq \mathbb{R}^n$, the *column space* of $\boldsymbol{A}$ as $\operatorname{col} \boldsymbol{A} = \operatorname{span}\{ \boldsymbol{c}_1, \cdots \boldsymbol{c}_n \} \sqsubseteq {}^m\mathbb{R}$
  - Both the row space and the column space have max dimension $\min\{ m, n \}$ because they're restricted by the number of vectors in the spanning set and the space it's a subspace of
- The column space of $\boldsymbol{A}$ is equal to its image: $\operatorname{col} \boldsymbol{A} = \{ \boldsymbol{y} \in {}^m\mathbb{R} \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \forall \boldsymbol{x} \in {}^n\mathbb{R} \}$
- Proposition I: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$ and $\boldsymbol{U} \in {}^m\mathbb{R}^m, \boldsymbol{V} \in {}^n\mathbb{R}^n$, then:
  1. $\operatorname{row} \boldsymbol{U}\boldsymbol{A} \sqsubseteq \operatorname{row} \boldsymbol{A}$
     - All the rows of $\boldsymbol{U}\boldsymbol{A}$ are linear combinations of the rows of $\boldsymbol{A}$
  2. $\operatorname{col} \boldsymbol{A}\boldsymbol{V} \sqsubseteq \operatorname{col} \boldsymbol{A}$
     - Similarly the columns of $\boldsymbol{A}\boldsymbol{V}$ are linear combinations of the columns of $\boldsymbol{A}$
  3. If $\boldsymbol{U}, \boldsymbol{V}$ are invertible, then $\operatorname{row} \boldsymbol{U}\boldsymbol{A} = \operatorname{row} \boldsymbol{A}$ and $\operatorname{col} \boldsymbol{A}\boldsymbol{V} = \operatorname{col} \boldsymbol{A}$
     - If $\boldsymbol{U}$ is invertible, consider $\boldsymbol{U} \to \boldsymbol{U}^{-1}$ and $\boldsymbol{A} \to \boldsymbol{U}\boldsymbol{A}$, so $\operatorname{row} \boldsymbol{U}\boldsymbol{A} \sqsubseteq \operatorname{row} \boldsymbol{A} \iff \operatorname{row} \boldsymbol{U}^{-1}(\boldsymbol{U}\boldsymbol{A}) \sqsubseteq \operatorname{row} \boldsymbol{U}\boldsymbol{A} \implies \operatorname{row} \boldsymbol{A} \sqsubseteq \operatorname{row} \boldsymbol{U}\boldsymbol{A}$
     - Since the two subspaces are within each other they must be equal
- Proposition II: Let $\{ \boldsymbol{x}_1, \cdots, \boldsymbol{x}_r \} \subset {}^m\mathbb{R}, \boldsymbol{U} \in {}^m\mathbb{R}^m$ invertible, then $\{ \boldsymbol{x}_1, \cdots, \boldsymbol{x}_r \}$ is linearly independent iff $\{ \boldsymbol{U}\boldsymbol{x}_1, \cdots, \boldsymbol{U}\boldsymbol{x}_r \}$ is linearly independent
  - Proof: $\sum_{i=1}^{n} \lambda_i(\boldsymbol{U}\boldsymbol{x}_i) = \boldsymbol{0} \iff \boldsymbol{U}\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i\right) = \boldsymbol{0} \iff \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i = \boldsymbol{0}$ so linearly independence of one set implies all $\lambda_i = 0$ which means the other set is linearly independent
  - We don't lose any information by multiplying a set of vectors by an invertible matrix

Lecture 14, Feb 14, 2022

Row Dimension Equals Column Dimension

- Lemma I: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$, then $\operatorname{row} \tilde{\boldsymbol{A}} = \operatorname{row} \boldsymbol{A}$ (where $\tilde{\boldsymbol{A}}$ is the row-reduced echelon form of $\boldsymbol{A}$), so $\dim \operatorname{row} \tilde{\boldsymbol{A}} = \dim \operatorname{row} \boldsymbol{A}$, and the nonzero rows of $\tilde{\boldsymbol{A}}$ form a basis for $\operatorname{row} \tilde{\boldsymbol{A}} = \operatorname{row} \boldsymbol{A}$
  - Proof:
    * $\tilde{\boldsymbol{A}} = \boldsymbol{E}_n \cdots \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{E}\boldsymbol{A}$; since E_i are elementary matrices they are invertible, therefore $\boldsymbol{E}$ is invertible, so by Prop. I, $\operatorname{row} \tilde{\boldsymbol{A}} = \operatorname{row} \boldsymbol{A}$
    * To show the rows of $\tilde{\boldsymbol{A}}$ form a basis, we show linear independence and span
      - The nonzero rows span $\operatorname{row} \tilde{\boldsymbol{A}} = \operatorname{row} \boldsymbol{A}$ because the zero rows add nothing to the span
      - The rows are linearly independent because there is only one nonzero entry in each column with a leading 1
- Lemma II: Let $A \in {}^m\mathbb{R}^n$, then (1) the columns with leading ones in $\tilde{\boldsymbol{A}}$ is a basis for $\operatorname{col} \tilde{\boldsymbol{A}}$, and (2) the columns in $\boldsymbol{A}$ *corresponding* to the columns with leading ones in $\tilde{\boldsymbol{A}}$ is a basis for $\operatorname{col} \boldsymbol{A}$
  - In other words, $\operatorname{col} \tilde{\boldsymbol{A}} \neq \operatorname{col} \boldsymbol{A}$, but we can get a basis for $\operatorname{col} \boldsymbol{A}$ by taking the columns in $\boldsymbol{A}$ that correspond to columns with leading ones in $\tilde{\boldsymbol{A}}$
  - As a result $\dim \operatorname{col} \tilde{\boldsymbol{A}} = \dim \operatorname{col} \boldsymbol{A}$
  - Proof (1):
    * Independence: The columns with leading ones in $\tilde{\boldsymbol{A}}$ are independent because they are a subset of the standard basis for ${}^m\mathbb{R}$
    * Generation: They also span $\tilde{\boldsymbol{A}}$ because the columns without leading ones can be expressed as a linear combination of the columns with leading ones

- Proof (2): ($\boldsymbol{c}_i$ denotes columns in $\boldsymbol{A}$, $\boldsymbol{c}'_i$ denotes columns in $\tilde{\boldsymbol{A}}$, $\boldsymbol{c}'_{ji}$ denotes column i with leading one in $\tilde{\boldsymbol{A}}$, $\boldsymbol{c}_{ji}$ denotes the column in $\boldsymbol{A}$ corresponding to column i with leading one in $\tilde{\boldsymbol{A}}$)
  * Independence: $\tilde{\boldsymbol{A}} = \boldsymbol{E}\boldsymbol{A} \implies \begin{bmatrix}\boldsymbol{c}'_1 & \cdots & \boldsymbol{c}'_n\end{bmatrix} = \begin{bmatrix}\boldsymbol{E}\boldsymbol{c}_1 & \cdots & \boldsymbol{E}\boldsymbol{c}_n\end{bmatrix} \implies \boldsymbol{c}'_i = \boldsymbol{E}\boldsymbol{c}_i \implies \boldsymbol{c}_i = \boldsymbol{E}^{-1}\boldsymbol{c}'_i$ so the set of columns in $\boldsymbol{A}$ corresponding to the leading ones columns in $\tilde{\boldsymbol{A}}$ are linearly independent as they are related by an invertible matrix by Prop. II
  * Generation: Let $\boldsymbol{y} = \sum_{i=1}^{n} \mu_i \boldsymbol{c}_i \in \operatorname{col} \boldsymbol{A}; y = \sum_{i=1}^{n} \mu_j (\boldsymbol{E}^{-1}\boldsymbol{c}'_j) = E^{-1}\left(\sum_{i=1}^{n} \mu_j \boldsymbol{c}'_j\right) \in \operatorname{col} \tilde{\boldsymbol{A}}$; but we've previously shown that the columns with leading ones form a basis for $\operatorname{col} \tilde{\boldsymbol{A}}$, so $E^{-1}\left(\sum_{i=1}^{n} \mu_j \boldsymbol{c}'_j\right) = E^{-1}\left(\sum_{i=1}^{n} \eta_i \boldsymbol{c}_{ji}\right) = \sum_{i=1}^{n} E^{-1}\eta_i \boldsymbol{c}'_{ji} = \sum_{i=1}^{n} \eta_i \boldsymbol{c}_{ji}$
- Lemma II gives us two methods of making a basis from a spanning set: making the vectors the rows of a matrix, reducing the matrix and then taking the nonzero rows as the basis, or making the vectors the columns, reducing the matrix and then taking the columns corresponding to columns with leading ones
- Theorem I: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$, then $\dim \operatorname{row} \boldsymbol{A} = \dim \operatorname{col} \boldsymbol{A}$
  - Proof: $\dim \operatorname{row} \boldsymbol{A} = \dim \operatorname{row} \tilde{\boldsymbol{A}} = r = \dim \operatorname{col} \tilde{\boldsymbol{A}} = \dim \operatorname{col} \boldsymbol{A}$

Lecture 15, Feb 15, 2022

Rank

- Definition: The *rank* of $\boldsymbol{A}$, denoted $\operatorname{rank} \boldsymbol{A}$, is the common dimension of its row and column space: $\operatorname{rank} \boldsymbol{A} \equiv \dim \operatorname{row} \boldsymbol{A} = \dim \operatorname{col} \boldsymbol{A}$
  - Can also be expressed in different ways, e.g. number of nonzero rows in the RREF, the number of leading ones in the RREF, etc
- Properties of rank:
  - Property I: $\operatorname{rank} \boldsymbol{A} = \operatorname{rank} \tilde{\boldsymbol{A}}$
  - Property II: $\operatorname{rank} \boldsymbol{A} = \operatorname{rank} \boldsymbol{A}^T$
  - Property III: $\operatorname{rank} \boldsymbol{U}\boldsymbol{A} \leq \operatorname{rank} \boldsymbol{A}$
    * $\operatorname{row} \boldsymbol{U}\boldsymbol{A} \subseteq \operatorname{row} \boldsymbol{A}$ by Prop. I
    * $\operatorname{rank} \boldsymbol{U}\boldsymbol{A} = \operatorname{rank} \boldsymbol{A}$ when $\boldsymbol{U}$ is invertible since $\operatorname{row} \boldsymbol{U}\boldsymbol{A} = \operatorname{row} \boldsymbol{A}$ by Prop. I
    * Similarly $\operatorname{rank} \boldsymbol{A}\boldsymbol{V} \leq \operatorname{rank} \boldsymbol{A}$ and $\operatorname{rank} \boldsymbol{A}\boldsymbol{V} = \operatorname{rank} \boldsymbol{A}$ if (but not only if) $\boldsymbol{V}$ is invertible

Lecture 16, Feb 18, 2022

The Dimension Formula

- Theorem II: The Dimension Formula: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$, then $\dim \operatorname{null} \boldsymbol{A} = n - \operatorname{rank} \boldsymbol{A}$ (or $\operatorname{rank} \boldsymbol{A} + \dim \operatorname{null} \boldsymbol{A} = n$)
  - Proof:
    * Let $\{ \boldsymbol{s}_1, \cdots, \boldsymbol{s}_k \}$ be a basis for $\operatorname{null} \boldsymbol{A}$; since $\operatorname{null} \boldsymbol{A} \sqsubseteq {}^n\mathbb{R}$, we can extend this to a basis for ${}^n\mathbb{R}$: $\{ \boldsymbol{s}_1, \cdots, \boldsymbol{s}_k, \boldsymbol{s}_{k+1}, \cdots, \boldsymbol{s}_n \}$
    * Claim: $\{ \boldsymbol{A}\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{A}\boldsymbol{s}_n \}$ is a basis for $\operatorname{col} \boldsymbol{A}$ (if this is true, then $\dim \operatorname{col} \boldsymbol{A} = n - k = n - \dim \operatorname{null} \boldsymbol{A}$ and we're done)

- Linear independence: $$\sum_{i=k+1}^{n} \lambda_i \boldsymbol{A}\boldsymbol{s}_i = \boldsymbol{0}$$ $$\implies \boldsymbol{A}\left(\sum_{i=k+1}^{n} \lambda_i \boldsymbol{s}_i\right) = \boldsymbol{0}$$ $$\implies \sum_{i=k+1}^{n} \lambda_i \boldsymbol{s}_i \in \operatorname{null} \boldsymbol{A}$$ $$\implies \sum_{i=k+1}^{n} \lambda_i \boldsymbol{s}_i = \sum_{i=1}^{k} \mu_i \boldsymbol{s}_i$$
  - Let $\mu_i = -\lambda_i \implies \sum_{i=1}^{n} \lambda_i \boldsymbol{s}_i = \boldsymbol{0} \implies \lambda_i = 0$ since $\{ s_i \}$ are a basis (note we can do this because i in the first summation and i in the second summation never have the same values)
- Generation: $\operatorname{span}\{ \boldsymbol{A}\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{A}\boldsymbol{s}_n \} \subseteq \operatorname{col} \boldsymbol{A}$ by Prop. I because each $\boldsymbol{s}_{k+1}, \cdots \boldsymbol{s}_n \in \operatorname{col} \boldsymbol{A}$
  - To go the other way: $$\boldsymbol{y} \in \operatorname{col} \boldsymbol{A}$$ $$\implies \exists \boldsymbol{x} \ni \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} = \sum_{i=1}^{n} \beta_i \boldsymbol{s}_i$$ $$\implies \boldsymbol{y} = \sum_{i=1}^{n} \beta_i \boldsymbol{A}\boldsymbol{s}_i$$ $$\implies \sum_{i=k+1}^{n} \beta_i \boldsymbol{A}\boldsymbol{s}_i$$ $$\implies \boldsymbol{y} \in \operatorname{span}\{ \boldsymbol{A}\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{A}\boldsymbol{s}_n \}$$
  - Therefore $\operatorname{col} A = \operatorname{span}\{ \boldsymbol{A}\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{A}\boldsymbol{s}_n \}$
- Consider $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$: there may exist no $\boldsymbol{x}$, or one unique $\boldsymbol{x}$, or infinitely many $\boldsymbol{x}$
  - No solution: $\boldsymbol{b} \notin \operatorname{col} \boldsymbol{A} \implies \operatorname{col} \boldsymbol{A} \subsetneq \operatorname{col}[\boldsymbol{A}|\boldsymbol{b}]$ or $\operatorname{rank} \boldsymbol{A} < \operatorname{rank}[\boldsymbol{A}|\boldsymbol{b}]$
  - Unique solution: $\boldsymbol{b} \in \operatorname{col} \boldsymbol{A}$ and $\operatorname{col} \boldsymbol{A} = \operatorname{col}[\boldsymbol{A}|\boldsymbol{b}]$ and $\operatorname{null} \boldsymbol{A} = \{ \boldsymbol{0} \}$
  - Infinite solutions: $\boldsymbol{b} \in \operatorname{col} \boldsymbol{A}$ and $\operatorname{col} \boldsymbol{A} = \operatorname{col}[\boldsymbol{A}|\boldsymbol{b}]$ and $\dim \operatorname{null} \boldsymbol{A} > 0$
- Theorem III: The following statements are equivalent for $\boldsymbol{A} \in {}^n\mathbb{R}^n$:
  1. $\boldsymbol{A}$ is invertible
  2. $\operatorname{rank} \boldsymbol{A} = n$ (i.e. $\boldsymbol{A}$ is full rank)
  3. $\boldsymbol{A}$ has linearly independent rows
  4. $\boldsymbol{A}$ has linearly independent columns
  5. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{0}$
  6. $\boldsymbol{z}^T\boldsymbol{A} = \boldsymbol{0}^T \implies \boldsymbol{z} = \boldsymbol{0}$
- Fredholm Alternative: Either $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has exactly one solution xor $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has a nontrivial solution
- Proof of Theorem III: In the case where we have a set of equivalent statements it's often most convenient to show a circular chain of implication, e.g. 1 implies 2 implies 3 implies 1
  - $1 \implies 2$: $\boldsymbol{A}$ is invertible means $\tilde{\boldsymbol{A}} = \boldsymbol{1} \implies \operatorname{rank} \boldsymbol{A} = \dim \operatorname{row} \boldsymbol{A} = \dim \operatorname{row} \tilde{\boldsymbol{A}} = n$
  - $2 \implies 3$: $\operatorname{rank} \boldsymbol{A} = n \implies \dim \operatorname{row} \boldsymbol{A} = n$ but $\boldsymbol{A}$ only has n rows, so they have to be linearly independent
  - $3 \implies 4$: Linearly independent rows $\implies \dim \operatorname{row} \boldsymbol{A} = n \implies \dim \operatorname{col} \boldsymbol{A} = n \implies$ the columns are linearly independent since there are n columns
  - $4 \implies 5$: Linearly independent columns $\implies \sum_{i=1}^{n} x_i \boldsymbol{c}_i = \boldsymbol{0} \implies x_i = 0 \implies \boldsymbol{x} = \boldsymbol{0}$
    * Alternatively the independent columns implies $\operatorname{rank} \boldsymbol{A} = n \implies \dim \operatorname{null} \boldsymbol{A} = 0 \implies \{ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{0} \}$
  - $5 \implies 6$: $\{ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{0} \} \implies \dim \operatorname{null} \boldsymbol{A} = 0 \implies \operatorname{rank} \boldsymbol{A} = n \implies \dim \operatorname{null} \boldsymbol{A}^T = 0 \implies \{ \boldsymbol{z}^T\boldsymbol{A} = \boldsymbol{0} \implies \boldsymbol{z} = \boldsymbol{0} \}$

- $6 \implies 1$: By contraposition, assume $\boldsymbol{A}$ is not invertible, which means there are zero rows in the rref, so $\operatorname{rank} \boldsymbol{A} < n$ so the rows are linearly dependent, which means $\boldsymbol{z}^T\boldsymbol{A} = \boldsymbol{0}$ has a nontrivial solution

Lecture 17, Feb 28, 2022

Non-Square Matrices

- Theorem IV: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$; then the following are equivalent:
  1. $\operatorname{rank} \boldsymbol{A} = n$
  2. The columns of $\boldsymbol{A}$ are linearly independent
  3. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{0}$
  4. $\boldsymbol{A}^T\boldsymbol{A}$ is invertible
  5. $\boldsymbol{A}$ has a left inverse ($\exists \boldsymbol{B} \ni \boldsymbol{B}\boldsymbol{A} = \boldsymbol{1} \in {}^n\mathbb{R}^n$), where $\boldsymbol{B} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is the left inverse, known as the Moore-Penrose pseudoinverse
- This necessitates $m \geq n$, i.e. $\boldsymbol{A}$ is a "tall" matrix, because if $n > m$ then the columns cannot be independent
- Proof:
  - $1 \implies 2$: $\operatorname{rank} \boldsymbol{A} = n \implies \dim \operatorname{col} \boldsymbol{A} = n$ so the columns are linearly independent as there are n columns
  - $2 \implies 3$: The columns are independent, so the only linear combination of the columns that add to 0 is all 0s, which is the zero vector
  - $3 \implies 4$: $\boldsymbol{A}^T\boldsymbol{A}$ is square, so it is invertible if and only if $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{0}$
    * Lemma III: Let $\boldsymbol{s} \in {}^n\mathbb{R}$ and $\boldsymbol{s}^T\boldsymbol{s} = 0$ then $\boldsymbol{s} = \boldsymbol{0}$
      - Proof: $\boldsymbol{s}^T\boldsymbol{s} = \sum_{i=1}^{n} s_i^2 = 0$ but each $s_i^2 \geq 0$, which means all $s_i^2 = 0 \implies s_i = 0 \implies \boldsymbol{s} = \boldsymbol{0}$
    * $$\begin{aligned} &\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \\ \implies &\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = 0 \\ \implies &(\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = 0 \\ \implies &\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \\ \implies &\boldsymbol{x} = \boldsymbol{0} \end{aligned}$$
  - $4 \implies 5$: $\boldsymbol{A}^T\boldsymbol{A}$ is invertible implies $\exists \boldsymbol{C} \ni \boldsymbol{C}\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{1}$; let $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{A}^T$, then $\boldsymbol{B}$ is the one-sided inverse
  - $5 \implies 1$: Show the columns are linearly independent: $$\begin{aligned} &\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{c}_i = \boldsymbol{0} \\ \implies &\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0} \\ \implies &\boldsymbol{B}\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0} \\ \implies &\boldsymbol{1}\boldsymbol{c} = \boldsymbol{0} \\ \implies &\boldsymbol{c} = \boldsymbol{0} \end{aligned}$$
- Theorem IV: Let $\boldsymbol{A} \in {}^m\mathbb{R}^n$ (this time $n \geq m$, i.e. $\boldsymbol{A}$ is short and wide); then the following are equivalent:
  1. $\operatorname{rank} \boldsymbol{A} = m$
  2. The rows of $\boldsymbol{A}$ are linearly independent
  3. $\boldsymbol{x}^T\boldsymbol{A} = \boldsymbol{0}^T \implies \boldsymbol{x} = \boldsymbol{0}$
  4. $\boldsymbol{A}\boldsymbol{A}^T$ is invertible
  5. $\boldsymbol{A}$ has a right inverse ($\exists \boldsymbol{B} \ni \boldsymbol{A}\boldsymbol{B} = \boldsymbol{1} \in {}^m\mathbb{R}^m$), where $\boldsymbol{B} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ (also the Moore-Penrose pseudoinverse)

Lecture 18, Mar 1, 2022

Linear Transformations/Operators

- Definition: A *linear transformation* is a transformation between two vector spaces $\mathcal{V}$ and $\mathcal{W}$, $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$ (where $\mathcal{V}$ is the *domain* and $\mathcal{W}$ is the *codomain*) that has the following properties:
  1. (L1) Distribution: $\mathscr{L}(\boldsymbol{u} + \boldsymbol{v}) = \mathscr{L}(\boldsymbol{u}) + \mathscr{L}(\boldsymbol{v}), \forall \boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}$
  2. (L2) Homogeneity: $\mathscr{L}(\lambda \boldsymbol{v}) = \lambda \mathscr{L}(\boldsymbol{v}), \forall \boldsymbol{v} \in \mathcal{V}, \lambda \in \Gamma$
- These properties can be combined into one as $\mathscr{L}(\lambda \boldsymbol{u} + \mu \boldsymbol{v}) = \lambda \mathscr{L}(\boldsymbol{u}) + \mu \mathscr{L}(\boldsymbol{v})$
- A matrix $\boldsymbol{A} \in {}^m\mathbb{R}^n$ can be thought of as a linear transformation of $\boldsymbol{A} : {}^m\mathbb{R} \mapsto {}^n\mathbb{R}$
- The trace $\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} a_{ii}$ is a linear transformation $\operatorname{tr} : {}^n\mathbb{R}^n \mapsto \mathbb{R}$
  - L1: $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \sum_{i=1}^{n} (a_{ii} + b_{ii}) = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = \operatorname{tr} \boldsymbol{A} + \operatorname{tr} \boldsymbol{B}$
  - L2: $\operatorname{tr}(\lambda \boldsymbol{A}) = \sum_{i=1}^{n} \lambda a_{ii} = \lambda \sum_{i=1}^{n} a_{ii} = \lambda \operatorname{tr} \boldsymbol{A}$
- Properties of linear transformations $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$:
  1. $\mathscr{L}(\boldsymbol{0}) = \boldsymbol{0}$
     - $\mathscr{L}(\boldsymbol{0}) = \mathscr{L}(0\boldsymbol{v}) = 0\mathscr{L}(\boldsymbol{v}) = \boldsymbol{0}$
  2. $\mathscr{L}(-\boldsymbol{v}) = -\mathscr{L}(\boldsymbol{v})$
     - $\mathscr{L}(-\boldsymbol{v}) = \mathscr{L}((-1)\boldsymbol{v}) = -1\mathscr{L}(\boldsymbol{v}) = -\mathscr{L}(\boldsymbol{v})$
  3. $\mathscr{L}\left(\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i\right) = \sum_{i=1}^{n} \lambda \mathscr{L}(\boldsymbol{v}_i)$
     - Proven by induction
- Definition: The *image* of a linear transformation $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$ is $\operatorname{im} \mathscr{L} = \{\, \boldsymbol{w} \in \mathcal{W} \mid \boldsymbol{w} = \mathscr{L}(\boldsymbol{v}), \forall \boldsymbol{v} \in \mathcal{V} \,\}$ (column space for a matrix)
  - The image is a subspace of $\mathcal{W}$
- $\mathscr{L}$ maps $\mathcal{V}$ *into* $\mathcal{W}$ if $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$ but $\operatorname{im} \mathscr{L} \neq \mathcal{W}$
- $\mathscr{L}$ maps $\mathcal{V}$ *onto* $\mathcal{W}$ if $\operatorname{im} \mathscr{L} = \mathcal{W}$ (*surjective*), i.e. $\forall \boldsymbol{w} \in \mathcal{W}, \exists \boldsymbol{v} \ni \mathscr{L}(\boldsymbol{v}) = \boldsymbol{w}$
- $\mathscr{L}$ is *injective* if it is one-to-one, i.e. no two vectors in $\mathcal{V}$ maps onto the same vector in $\mathcal{W}$: $\nexists \boldsymbol{v}_1 = \boldsymbol{v}_2 \in \mathcal{V} \ni \mathscr{L}(\boldsymbol{v}_1) = \mathscr{L}(\boldsymbol{v}_2)$
- If $\mathscr{L}$ is surjective and injective then it is *bijective*
  - Bijective transformations have inverses
- Definition: The *kernel* of a linear transformation $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$ is $\ker \mathscr{L} = \{\, \boldsymbol{v} \in \mathcal{V} \mid \mathscr{L}(\boldsymbol{v}) = \boldsymbol{0} \,\}$ (null space for a matrix)
  - The kernel is a subspace of $\mathcal{V}$
  - The kernel is everything in $\mathcal{V}$ that maps to $\boldsymbol{0}$ in $\mathcal{W}$
- The dimension formula for linear transformations: $\dim \ker \mathscr{L} + \dim \operatorname{im} \mathscr{L} = \dim \mathcal{V}$
  - Analogous to the dimension formula for matrices; for matrices $\ker \boldsymbol{A} = \operatorname{null} \boldsymbol{A}$, $\operatorname{im} \boldsymbol{A} = \operatorname{col} \boldsymbol{A}$ and $\dim \mathcal{V} = n$

Lecture 19, Mar 4, 2022

Linear Transformations and Matrices

- Not only do all matrices represent linear transformations, all linear transformations can be represented as a matrix; there exists a one-to-one relationship between matrices and linear transformations
- Consider $\mathscr{L} : \mathcal{V} \mapsto \mathcal{W}$ and $\boldsymbol{w} \in \mathcal{W} = \mathscr{L}(\boldsymbol{v} \in \mathcal{V})$
  - $\boldsymbol{v} = \sum_{j=1}^{n} v_j \boldsymbol{e}_j$ given $E = \{\, \boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \,\}$ is a basis for $\mathcal{V}$

- $\boldsymbol{w} = \mathscr{L}\left(\sum_{j=1}^{n} v_j \boldsymbol{e}_j\right) = \sum_{j=1}^{n} v_j \mathscr{L}(\boldsymbol{e}_j)$
- $\boldsymbol{w} = \sum_{i=1}^{m} w_i \boldsymbol{h}_i$ where $H = \{ \boldsymbol{h}_1, \cdots, \boldsymbol{h}_m \}$ is a basis for $\mathcal{W}$
- $\mathscr{L}(\boldsymbol{e}_j) = \sum_{i=1}^{m} l_{ij} \boldsymbol{h}_i \implies \boldsymbol{w} = \sum_{j=1}^{n} v_j \mathscr{L}(\boldsymbol{e}_j) = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{m} l_{ij} \boldsymbol{h}_i\right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} l_{ij} v_j\right) \boldsymbol{h}_i$
- Compare the 2 lines above, we get $w_i = \sum_{j=1}^{n} l_{ij} v_j$ which is a matrix multiplication: $\boldsymbol{w} = \boldsymbol{L}\boldsymbol{v}$
  - $\boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}, \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \boldsymbol{L} = \begin{bmatrix} l_{ij} \end{bmatrix}$
- The v_j are *coordinates* of $\boldsymbol{v}$ with respect to the basis E; w_i are coordinates of $\boldsymbol{w}$ with respect to the basis H

- There is a one-to-one relationship between $\boldsymbol{w} = \mathscr{L}(\boldsymbol{v})$ and $\boldsymbol{w} = \boldsymbol{L}\boldsymbol{v}$

Lecture 20, Mar 7, 2022

Change of Basis

- Say we have a vector expressed in a vector space as coordinates with respect to one set of basis vectors; how do we express them in terms of another basis?
- The new coordinates in terms of the new basis is related to the old coordinates by a matrix
- $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{v}'$ where $\boldsymbol{P}$ is the *transition matrix* or *transformation matrix* or *change-of-basis matrix*
- Let $\mathcal{V}$ where $\dim \mathcal{V} = n$ have 2 bases: $E = \{ \boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \}$ and $F = \{ \boldsymbol{f}_1, \cdots, \boldsymbol{f}_n \}$
  - $\boldsymbol{v} = \sum_{i=1}^{n} v_i^{(e)} \boldsymbol{e}_i = \sum_{i=1}^{n} v_i^{(f)} \boldsymbol{f}_i$
  - Let $\boldsymbol{v}_e = \begin{bmatrix} v_1^{(e)} \\ \vdots \\ v_n^{(e)} \end{bmatrix}, \boldsymbol{v}_f = \begin{bmatrix} v_1^{(f)} \\ \vdots \\ v_n^{(f)} \end{bmatrix}$
    - $\boldsymbol{v}_e$ are the coordinates in terms of E, $\boldsymbol{v}_f$ are the coordinates in terms of F
  - What is the relationship between $\boldsymbol{v}_e$ and $\boldsymbol{v}_f$?

- Since the bases live in $\mathcal{V}$, in general $\boldsymbol{e}_j = \sum_{i=1}^{n} p_{ij} \boldsymbol{f}_i$
  - $$\begin{aligned} \boldsymbol{v} &= \sum_{i=1}^{n} v_i^{(f)} \boldsymbol{f}_i \\ &= \sum_{j=1}^{n} v_j^{(e)} \boldsymbol{e}_j \\ &= \sum_{j=1}^{n} v_j^{(e)} \left(\sum_{i=1}^{n} p_{ij} \boldsymbol{f}_i\right) \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{ij} v_j^{(e)}\right) \boldsymbol{f}_i \end{aligned}$$
  - $v_i^{(f)} = \sum_{j=1}^{n} p_{ij} v_j^{(e)}$ or $\boldsymbol{v}_f = \boldsymbol{P}\boldsymbol{v}_e$
  - The *columns* of $\boldsymbol{P}$ are the basis $\boldsymbol{e}_n$ expressed in terms of $\boldsymbol{f}_n$

- Proposition III: Let $B_e = \{ \boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \}$ be the standard basis for ${}^n\mathbb{R}$ and $B_f = \{ \boldsymbol{f}_1, \cdots \boldsymbol{f}_n \}$ be another basis; then the transition matrix from B_f to B_e is $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{f}_1 & \cdots & \boldsymbol{f}_n \end{bmatrix}$
- There is an isomorphism between $\mathcal{V} \leftrightarrow {}^n\mathbb{R}$; instead of thinking of members of $\mathcal{V}$ directly, we can think about their coordinates, which are vectors in ${}^n\mathbb{R}$
- Theorem II: Let $B_e = \{ \boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \} \subset \mathcal{V}$ be a basis for $\mathcal{V}$ and the coordinates $\boldsymbol{v}' \in {}^n\mathbb{R}$ for $\boldsymbol{v} \in \mathcal{V}$; then $\{ \boldsymbol{v}_1, \cdots, \boldsymbol{v}_m \}$ is linearly independent iff the coordinates $\{ \boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_m \}$ are linearly independent
  - Note notation $\boldsymbol{v} = v_1\boldsymbol{e}_1 + \cdots + v_n\boldsymbol{e}_n = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \boldsymbol{\mathscr{E}}\boldsymbol{v}$ where $\boldsymbol{\mathscr{E}} \in \mathcal{V}^n$, formalized as $\mathcal{V}^n \times {}^n\mathbb{R} \mapsto \mathcal{V}$
  - Proposition IV:
    * If $\boldsymbol{\mathscr{E}}\boldsymbol{v} = \boldsymbol{0}$ then $\boldsymbol{v} = \boldsymbol{0}$ because $\{ \boldsymbol{e}_1, \cdots, \boldsymbol{e}_m \}$ are linearly independent
    * If $\boldsymbol{\mathscr{E}}\boldsymbol{v} = \boldsymbol{\mathscr{E}}\boldsymbol{u}$ then $\boldsymbol{v} = \boldsymbol{u}$ since there is only one way to express a given vector as a linear combination of a set of independent vectors
  - Proof: $\displaystyle\sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j = \boldsymbol{0} \iff \sum_{j=1}^{n} \lambda_j(\boldsymbol{\mathscr{E}}\boldsymbol{v}'_j) = \boldsymbol{0} \iff \boldsymbol{\mathscr{E}}\left(\sum_{j=1}^{n} = \lambda_j \boldsymbol{v}'_j\right) = \boldsymbol{0} \iff \sum_{j=1}^{n} = \lambda_j \boldsymbol{v}'_j = \boldsymbol{0}$

Lecture 21, Mar 8, 2022

The Determinant Function

- Every matrix has a determinant denoted $\det(\boldsymbol{A})$; for 2×2 matrix this is $ad - bc$
- Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \\ \boldsymbol{r}_n \end{bmatrix} \in {}^n\mathbb{R}^n$; then the *determinant function* $\Delta_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ is any function that satisfies the following:
  1. Adding one row to another row leaves the result unchanged: $\Delta_n \left[\boldsymbol{E}(1; i, j)\boldsymbol{A}\right] = \Delta_n(\boldsymbol{A})$
     - $\boldsymbol{E}(\lambda; i, j)$ is an elementary matrix of type III that multiplies row j by λ and adds it to row i
  2. $\Delta_n \left[\boldsymbol{E}(\lambda, i)\boldsymbol{A}\right] = \lambda\Delta_n(\boldsymbol{A})$
     - $\boldsymbol{E}(\lambda, i)$ is an elementary matrix of type II that multiplies row i by λ
- The determinant function is homogeneous in each row (i.e. scaling a row scales the entire determinant)
- Theorem I: $\Delta_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ has the properties:
  1. If $\boldsymbol{A}$ has a zero row, then $\Delta_n(\boldsymbol{A}) = 0$
     - Proof: If row i of $\boldsymbol{A}$ is zero, then $\boldsymbol{E}(0, i)\boldsymbol{A} = \boldsymbol{A}$, therefore $\Delta_n \left[\boldsymbol{E}(0, i)\boldsymbol{A}\right] = \Delta_n(\boldsymbol{A}) = 0\Delta_n(\boldsymbol{A}) = 0$
  2. $\Delta_n \left[\boldsymbol{E}(\lambda; i, j)\boldsymbol{A}\right] = \Delta_n(\boldsymbol{A})$ (property 1, but with any scalar multiple)
     - Proof: Trivially true for $\lambda = 0$; for nonzero λ, scale row j by λ (scales the determinant by j), then add row j to row i (determinant unchanged), then divide the result by λ (determent scaled by $\frac{1}{\lambda}$), which gives the same determinant
  3. Interchanging rows negates the determinant: $\Delta_n \left[\boldsymbol{E}(i, j)\boldsymbol{A}\right] = -\Delta_n(\boldsymbol{A})$
     - Proof: $\Delta_n \begin{bmatrix} \vdots \\ \boldsymbol{r}_j \\ \vdots \\ \boldsymbol{r}_i \\ \vdots \end{bmatrix} = -\Delta_n \begin{bmatrix} \vdots \\ -\boldsymbol{r}_j \\ \vdots \\ \boldsymbol{r}_i \\ \vdots \end{bmatrix} = -\Delta_n \begin{bmatrix} \vdots \\ -\boldsymbol{r}_j \\ \vdots \\ \boldsymbol{r}_i - \boldsymbol{r}_j \\ \vdots \end{bmatrix} = \Delta_n \begin{bmatrix} \vdots \\ -\boldsymbol{r}_j \\ \vdots \\ \boldsymbol{r}_j - \boldsymbol{r}_i \\ \vdots \end{bmatrix} = \Delta_n \begin{bmatrix} \vdots \\ -\boldsymbol{r}_i \\ \vdots \\ \boldsymbol{r}_j - \boldsymbol{r}_i \\ \vdots \end{bmatrix} =$

$$-\Delta_n \begin{bmatrix} \vdots \\ \boldsymbol{r}_i \\ \vdots \\ \boldsymbol{r}_j - \boldsymbol{r}_i \\ \vdots \end{bmatrix} = -\Delta_n \begin{bmatrix} \vdots \\ \boldsymbol{r}_i \\ \vdots \\ \boldsymbol{r}_j \\ \vdots \end{bmatrix}$$

4. If the rows of $\boldsymbol{A}$ are linearly dependent then $\Delta_n(\boldsymbol{A}) = 0$
   - Proof: If the rows are dependent, then at least one row can be written as a linear combination of the others; therefore by adding multiples of other rows to this row (does not change the determinant by property 2), it is possible to make this row all zero, which means the determinant is 0 (by property 1)
5. The determinant function is linear in every row (n-linear): $\Delta_n \begin{bmatrix} \vdots \\ \lambda\boldsymbol{p} + \mu\boldsymbol{q} \\ \vdots \end{bmatrix} = \lambda\Delta_n \begin{bmatrix} \vdots \\ \boldsymbol{p} \\ \vdots \end{bmatrix} + \mu\Delta_n \begin{bmatrix} \vdots \\ \boldsymbol{q} \\ \vdots \end{bmatrix}$
   - Proof: Without loss of generality, show $\Delta_n \begin{bmatrix} \boldsymbol{p} + \boldsymbol{q} \\ \vdots \end{bmatrix} = \Delta_n \begin{bmatrix} \boldsymbol{p} \\ \vdots \end{bmatrix} + \Delta_n \begin{bmatrix} \boldsymbol{q} \\ \vdots \end{bmatrix}$
     * If the rest of the rows are dependent, then by property 4 each determinant is 0 so $0 = 0 + 0$
     * If the rest of the rows are independent, extend the rest of the rows to a basis for $\mathbb{R}^n$ by adding an independent vector; then $\boldsymbol{p} = \sum_{i=1}^{n} \lambda_i \boldsymbol{r}_i$ and $\boldsymbol{q} = \sum_{i=1}^{n} \mu_k \boldsymbol{r}_k$ so $\Delta_n \begin{bmatrix} \boldsymbol{p} + \boldsymbol{q} \\ \vdots \end{bmatrix} = \Delta_n \begin{bmatrix} (\lambda_1 + \mu_1)\boldsymbol{r}_1 + \sum_{k=2}^{n} (\lambda_k + \mu_k)\boldsymbol{r}_k \\ \vdots \end{bmatrix} = \Delta_n \begin{bmatrix} (\lambda_1 + \mu_1)\boldsymbol{r}_1 \\ \vdots \end{bmatrix} = (\lambda_1 + \mu_1)\Delta_n \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \end{bmatrix}$; also by the same process $\Delta_n \begin{bmatrix} \boldsymbol{p} \\ \vdots \end{bmatrix} = \lambda_1 \Delta_n \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \end{bmatrix}$ and $\Delta_n \begin{bmatrix} \boldsymbol{q} \\ \vdots \end{bmatrix} = \mu_1 \Delta_n \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \end{bmatrix}$

Lecture 22, Mar 11, 2022

More Determinant Properties

- Proposition I: Let $\boldsymbol{D} \in {}^n\mathbb{R}^n$ be a diagonal matrix with entries $d_1, \cdots, d_n$ on the diagonal, then $\Delta_n(\boldsymbol{D}) = \Delta_n(\boldsymbol{1}) \prod_{i=1}^{n} d_n$
- Proposition II: Let $\boldsymbol{U} \in {}^n\mathbb{R}^n$ be an upper triangular matrix, then $\Delta_n(\boldsymbol{U}) = \Delta_n(\boldsymbol{1}) \prod_{i=1}^{n} u_{ii}$
  - Start at the bottom row, make it a 1, cancel the column, etc

Lecture 23, Mar 14, 2022

The Determinant

- Lemma I: If $\Delta_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ is a determinant function, then $\Delta_n(\boldsymbol{A}) = \kappa(\boldsymbol{A})\Delta_n(\boldsymbol{1})$ where $\kappa(\boldsymbol{A})$ is a scalar function of $\boldsymbol{A}$
  - Proof: Gaussian eliminate on $\boldsymbol{A}$ until it is upper triangular, i.e. $\boldsymbol{E}_1 \cdots \boldsymbol{E}_n \boldsymbol{U}$; $\Delta_n(\boldsymbol{U}) = \Delta_n(\boldsymbol{1}) \prod_{i=1}^{n} u_{ii}$ which is a scalar times $\Delta_n(\boldsymbol{1})$; since elementary matrices either scale, negate, or leave the determinant unchanged, the final result is going to be a scalar times $\Delta_n(\boldsymbol{1})$

- Theorem II: Let $\Delta_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ and $\hat{\Delta}_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$; $\hat{\Delta}_n$ satisfies an additional property DIII $\hat{\Delta}_n(\mathbf{1}) = 1$; then $\Delta_n(\boldsymbol{A}) = \hat{\Delta}_n(\boldsymbol{A})\Delta_n(\mathbf{1})$
  - Proof: Consider $\Delta_n(\boldsymbol{A}) - \hat{\Delta}_n(\boldsymbol{A})\Delta_n(\mathbf{1}) = \kappa(\boldsymbol{A})\Delta_n(\mathbf{1}) - \kappa(\boldsymbol{A})\hat{\Delta}_n(\mathbf{1})\Delta_n(\mathbf{1})$
    $= \kappa(\boldsymbol{A})\Delta_n(\mathbf{1}) - \kappa(\boldsymbol{A})\Delta_n(\mathbf{1})$
    $= 0$
  - Corollary: If $\Delta_n(\mathbf{1}) = 1$ as well, then $\Delta_n(\boldsymbol{A}) = \hat{\Delta}_n(\boldsymbol{A})$
    * If DIII also holds, then the determinant function is unique
- Definition: The *determinant* of $\boldsymbol{A} \in {}^n\mathbb{R}^n$ is the unique determinant function $\Delta_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ that satisfies:
  1. DI: Adding one row to another row leaves the result unchanged: $\Delta_n\left[\boldsymbol{E}(1;i,j)\boldsymbol{A}\right] = \Delta_n(\boldsymbol{A})$
     - $\boldsymbol{E}(\lambda;i,j)$ is an elementary matrix of type III that multiplies row j by λ and adds it to row i
  2. DII: $\Delta_n\left[\boldsymbol{E}(\lambda,i)\boldsymbol{A}\right] = \lambda\Delta_n(\boldsymbol{A})$
     - $\boldsymbol{E}(\lambda,i)$ is an elementary matrix of type II that multiplies row i by λ
  3. DIII: $\Delta_n(\mathbf{1}) = 1$
- Definition: The (i,j) *minor* of a square matrix $\boldsymbol{A} \in {}^n\mathbb{R}^n$, denoted $\boldsymbol{M}_{ij}(\boldsymbol{A}) \in {}^{n-1}\mathbb{R}^{n-1}$, is the matrix obtained by eliminating the i-th row and j-th column
- Definition: The function $\det_n : {}^n\mathbb{R}^n \mapsto \mathbb{R}$ is $\det_n\boldsymbol{A} = \displaystyle\sum_{k=1}^{n}(-1)^{k+j}a_{kj}\det_{n-1}\boldsymbol{M}_{kj}(\boldsymbol{A})$ for any $1 \leq j \leq n$ and $\det_1\left[a\right] = a$
  - We can use any column and the definition still works
  - This is known as the *Laplace expansion*
- Theorem III: $\det_n {}^n\mathbb{R}^n \mapsto \mathbb{R}$ is the determinant
  - This shows the existence and uniqueness of the determinant
- The determinant can also be denoted $|\boldsymbol{A}|$

Lecture 24, Mar 15, 2022

Additional Properties of the Determinant

- Determinant of elementary matrices: $\det \boldsymbol{E}(i,j) = -1$, $\det \boldsymbol{E}(\lambda;i) = \lambda$, $\det \boldsymbol{E}(\lambda;i,j) = 1$
- Theorem IV: Cauchy-Binet Product Rule: Let $\boldsymbol{A},\boldsymbol{B} \in {}^n\mathbb{R}^n$, then $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
  - Proof: Define $\Delta_{\boldsymbol{B}}(\boldsymbol{A}) = \det(\boldsymbol{AB})$
    * Claim: $\Delta_{\boldsymbol{B}}$ is a proper determinant function:
      1. $\Delta_{\boldsymbol{B}}\left[\boldsymbol{E}(1;i,j)\boldsymbol{A}\right] = \det\left[\boldsymbol{E}(1;i,j)\boldsymbol{AB}\right] = \det(\boldsymbol{AB}) = \Delta_{\boldsymbol{B}}(\boldsymbol{A})$
      2. $\Delta_{\boldsymbol{B}}\left[\boldsymbol{E}(\lambda;i)\boldsymbol{A}\right] = \det\left[\boldsymbol{E}(\lambda;i)\boldsymbol{AB}\right] = \lambda\det(\boldsymbol{AB}) = \lambda\Delta_{\boldsymbol{B}}(\boldsymbol{A})$
    * From $\Delta_n(\boldsymbol{A}) = \det(\boldsymbol{A})\Delta_n(\mathbf{1})$ we know $\Delta_{\boldsymbol{B}}(\boldsymbol{A}) = \det(\boldsymbol{A})\Delta_{\boldsymbol{B}}(\mathbf{1}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
    * Therefore $\det(\boldsymbol{A})\det(\boldsymbol{B}) = \Delta_{\boldsymbol{B}}(\boldsymbol{A}) = \det(\boldsymbol{AB})$
- Theorem V: Transpose Rule: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$, then $\det\boldsymbol{A} = \det\boldsymbol{A}^T$
  - This means we can also compute the determinant along rows instead of columns, since the matrix can be transposed and the determinant is unchanged
  - Proof: Define $\Delta_T(\boldsymbol{A}) = \det\boldsymbol{A}^T$
    * Claim: Δ_T is a proper determinant function:
      1. $\Delta_T\left[\boldsymbol{E}(1;i,j)\boldsymbol{A}\right] = \det(\boldsymbol{EA})^T = \det(\boldsymbol{A}^T\boldsymbol{E}^T) = \det(\boldsymbol{A}^T)\det(\boldsymbol{E}^T) = \det(\boldsymbol{A}^T)\det(\boldsymbol{E}(\lambda;j,i)) = \det(\boldsymbol{A}^T) = \Delta_T(\boldsymbol{A})$
      2. $\Delta_T\left[\boldsymbol{E}(\lambda;i)\boldsymbol{A}\right] = \det(\boldsymbol{A}^T)\det(\boldsymbol{E}^T) = \det(\boldsymbol{A}^T)\det(\boldsymbol{E}(\lambda;i)) = \lambda\det(\boldsymbol{A}^T) = \lambda\Delta_T(\boldsymbol{A})$
      3. $\Delta_T(\mathbf{1}) = \det(\mathbf{1}^T) = \det\mathbf{1} = 1$
    * Therefore Δ_T is the determinant, and since the determinant is unique, $\det A^T = \Delta_T(\boldsymbol{A}) = \det\boldsymbol{A}$
- Theorem VI: Invertibility theorem: $\boldsymbol{A} \in {}^n\mathbb{R}^n$ is invertible iff $\det\boldsymbol{A} \neq 0$
  - Proof:
    * If $\boldsymbol{A}$ invertible, then $\boldsymbol{AA}^{-1} = \mathbf{1} \implies \det(\boldsymbol{AA}^{-1}) = \det\mathbf{1} = 1 \implies \det(\boldsymbol{A})\det(\boldsymbol{A}^{-1}) = 1$ so $\det(\boldsymbol{A}) \neq 0$

- Corollary: $\det \boldsymbol{A}^{-1} = \dfrac{1}{\det \boldsymbol{A}}$ if $\boldsymbol{A}$ is invertible
  * By contraposition, if $\boldsymbol{A}$ is not invertible, then its rows are dependent, then $\det \boldsymbol{A} = 0$; therefore $\det \boldsymbol{A} \neq 0 \implies \boldsymbol{A}$ is invertible

Lecture 25, Mar 18, 2022

Cramer's Rule

- Cramer's Rule (Maclaurin-Cramer Rule): The solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where $\boldsymbol{A} \in {}^n\mathbb{R}^n$ is given by $x_i = \dfrac{\det \boldsymbol{A}_i}{\det \boldsymbol{A}}$ where $\boldsymbol{x}_i$ are the components of $\boldsymbol{x}$ and $\boldsymbol{A}_i$ is $\boldsymbol{A}$ with column i replaced by $\boldsymbol{b}$, if $\det \boldsymbol{A} \neq 0$
  - $\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{b} & \cdots & \boldsymbol{c}_n \end{bmatrix}$
  - $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_n \end{bmatrix} \boldsymbol{x} = \displaystyle\sum_{j=1}^{n} x_j \boldsymbol{c}_j$
  - $$\begin{aligned} \det \boldsymbol{A}_i &= \det \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \displaystyle\sum_{j=1}^{n} x_j \boldsymbol{c}_j & \cdots & \boldsymbol{c}_n \end{bmatrix} \\ &= \det \begin{bmatrix} \boldsymbol{c}_1 & \cdots & x_i \boldsymbol{c}_i + \displaystyle\sum_{\substack{j=1 \\ j \neq i}}^{n} x_j \boldsymbol{c}_j & \cdots & \boldsymbol{c}_n \end{bmatrix} \\ &= \det \begin{bmatrix} \boldsymbol{c}_1 & \cdots & x_i \boldsymbol{c}_i & \cdots & \boldsymbol{c}_n \end{bmatrix} \\ &= x_i \det \begin{bmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_i & \cdots & \boldsymbol{c}_n \end{bmatrix} \\ &= x_i \det \boldsymbol{A} \end{aligned}$$
  - Provided that $\det \boldsymbol{A} \neq 0$, we have $\det \boldsymbol{A}_i = x_i \det \boldsymbol{A} \implies x_i = \dfrac{\det \boldsymbol{A}_i}{\det \boldsymbol{A}}$
  - Note the sum is just adding multiples of other columns, which does not change the determinant as per determinant properties
- Cramer's rule is computationally inefficient for larger matrices ($\mathcal{O}((n+1)!)$ operations if taking determinants recursively); Gaussian elimination is much better for larger matrices ($\mathcal{O}(n^3)$ operations)
  - For smaller matrices it might be faster; in general this depends on the matrix itself (e.g. number of zeros)

Cofactors and Adjoints

- Definition: The (i, j) cofactor of $\boldsymbol{A} \in {}^n\mathbb{R}^n$ is $c_{ij}(\boldsymbol{A}) = (-1)^{i+j} \det \boldsymbol{M}_{ij}(\boldsymbol{A})$
  - Using the cofactor, the determinant can be written as $\displaystyle\sum_{j=1}^{n} a_{kj} c_{kj}$, true for any k
- $\displaystyle\sum_{j=1}^{n} a_{ij} c_{kj} = \begin{cases} \det \boldsymbol{A} & k = i \\ 0 & k \neq i \end{cases}$
  - Proof: Consider $\boldsymbol{A}' \in {}^n\mathbb{R}^n$ which is $\boldsymbol{A}$ with row k replaced with row i
    * $\det \boldsymbol{A}' = 0$ because rows are not independent
    * Using the Laplace expansion about row k: $\det \boldsymbol{A}' = 0 = \displaystyle\sum_{j=1}^{n} a'_{kj} c_{kj}(\boldsymbol{A}') = \sum_{j=1}^{n} a_{ij} c_{kj}(\boldsymbol{A})$
      - $a'_{kj} = a_{ij}$ because we replaced row k by row i
      - $c_{kj}(\boldsymbol{A}') = c_{kj}(\boldsymbol{A})$ because row k was eliminated in the calculation of the cofactor so the minors are the same
- $\displaystyle\sum_{j=1}^{n} a_{ij} c_{kj}$ is like taking $\boldsymbol{A}\boldsymbol{C}^T$ where $\boldsymbol{C} = \begin{bmatrix} c_{kj} \end{bmatrix}$ is the cofactor matrix

- $\begin{cases} \det \boldsymbol{A} & k = i \\ 0 & k \neq i \end{cases}$ is just $(\det \boldsymbol{A})\mathbf{1}$
- $\boldsymbol{A}\boldsymbol{C}^T = (\det \boldsymbol{A})\mathbf{1}$
- Definition: The *adjoint* of $\boldsymbol{A}$ is $\operatorname{adj} \boldsymbol{A} = \boldsymbol{C}^T$
  - Also known as the *adjugate*
- Theorem VIII: $\boldsymbol{A}(\operatorname{adj} \boldsymbol{A}) = (\det \boldsymbol{A})\mathbf{1} = (\operatorname{adj} \boldsymbol{A})\boldsymbol{A}$
  - If $\boldsymbol{A}$ is invertible then $\boldsymbol{A}^{-1} = \dfrac{1}{\det \boldsymbol{A}} \operatorname{adj} \boldsymbol{A}$
- $\det(\operatorname{adj} \boldsymbol{A}) = (\det \boldsymbol{A})^{n-1}$
- If $\boldsymbol{A}$ is non-invertible, then $(\operatorname{adj} \boldsymbol{A})\boldsymbol{A} = \mathbf{0} \implies \operatorname{col} \boldsymbol{A} \subseteq \operatorname{null} \operatorname{adj} \boldsymbol{A}$; if $\boldsymbol{A} \neq \mathbf{0}$, $\dim \operatorname{col} \boldsymbol{A} \geq 1 \implies \dim \operatorname{null} \operatorname{adj} \boldsymbol{A} \geq 1$
  - $n - \operatorname{rank} \operatorname{adj} \boldsymbol{A} = \dim \operatorname{null} \operatorname{adj} \boldsymbol{A} \implies \operatorname{rank} \operatorname{adj} \boldsymbol{A} < n \implies \operatorname{adj} \boldsymbol{A}$ is not invertible
  - $\det(\operatorname{adj} \boldsymbol{A}) = 0$ so the previous equation still holds

Lecture 27, Mar 22, 2022

Eigenvalues and Eigenvectors: Definition and Motivation

- Motivation: Finding solutions to a system of differential equations $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} = \boldsymbol{x}(t) \in {}^n\mathbb{R}$, where the dot indicates time derivative
  - Assume that $\boldsymbol{A} \in {}^n\mathbb{R}^n$ is constant
  - Each equation is first order, but higher order equations can also be expressed in this form by making derivatives also variables
- As in the case for scalars, try $\boldsymbol{x}(t) = \boldsymbol{p}e^{\lambda t}$ where $\boldsymbol{p} \in {}^n\mathbb{R}$
  - $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \implies \lambda \boldsymbol{p} e^{\lambda t} = \boldsymbol{A}\boldsymbol{p}e^{\lambda t} \implies \lambda p = \boldsymbol{A}\boldsymbol{p} \implies (\lambda \mathbf{1} - \boldsymbol{A})\boldsymbol{p} = \mathbf{0}$
    - This is similar to the characteristic equation in the scalar case
    - We can say that $\boldsymbol{p} \neq \mathbf{0}$ since that would give the trivial solution
    - This means that $(\lambda\mathbf{1} - \boldsymbol{A})$ must have a null space, which means $\lambda\mathbf{1} - \boldsymbol{A}$ cannot have full rank, so we must choose λ such that $(\lambda\mathbf{1} - \boldsymbol{A})$ is singular, i.e. $\det(\lambda\mathbf{1} - \lambda A) = 0$
    - The "eigenproblem"
  - The λ that make $\det(\lambda\mathbf{1} - \lambda\boldsymbol{A}) = 0$ are the *eigenvalues* of $\boldsymbol{A}$
  - The nontrivial $\boldsymbol{p}$ are the *eigenvectors* (for a particular λ)
    - Note these can be scaled arbitrarily
- For $\boldsymbol{A} \in {}^n\mathbb{R}^n$, there are n such λ, because $\det(\lambda\mathbf{1} - \lambda\boldsymbol{A})$ is an n-th degree polynomial of λ
  - Expanding out the determinant, we obtain the *characteristic polynomial* (eigenpolynomial?) of this system of differential equations; when we set it to zero, we obtain the *characteristic equation* (eigenequation?)
  - Notation: $C_{\boldsymbol{A}}(\lambda)$ for the eigenpolynomial
- Since $\boldsymbol{p} \in \operatorname{null}(\lambda\mathbf{1} - \boldsymbol{A})$, the *eigenspace* for an eigenvalue λ is $\{\, \boldsymbol{p} \in {}^n\mathbb{R} \mid \boldsymbol{A}\boldsymbol{p} = \lambda\boldsymbol{p} \,\} = \operatorname{null}(\lambda\mathbf{1} - \lambda\boldsymbol{A})$ (sometimes denoted $\mathscr{E}_\lambda$)
  - The bases for the eigenspaces are the eigenvectors
  - All the eigenvectors are linearly independent
  - Note that since $\mathbf{0}$ is the trivial eigenvector, normally we use "eigenvector" to refer to only nonzero eigenvectors
- If $\boldsymbol{A}$ is viewed as a linear transformation, eigenvectors are the vectors that are scaled by the transformation by an eigenvalue (i.e. direction remains unchanged)
- This allows us to solve the general n-th order differential equation
  - Let $x_1 = x, x_2 = \dot{x}$, then $\dot{x}_1 = x_2, \dot{x}_2 = \ddot{x} = -a_1\dot{x} - a_0 x = -a_1 x_2 - a_0 x_1$
  - We can put this in a matrix as $\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
  - By extension this can be used to solve a linear system of any order
- The eigenvalues of an upper triangular matrix are the values on the diagonal of the matrix (since the determinant of such a matrix is the product of the diagonal)
- "Eigen" is a German word meaning "characteristic, proper"

Lecture 28, Mar 25, 2022

Properties of Eigenvalues and Eigenvectors

- If $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the eigenvalues satisfy $\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 - (\operatorname{tr}\boldsymbol{A})\lambda + \det\boldsymbol{A} = 0$
- In general, for $\boldsymbol{A} \in {}^n\mathbb{R}^n$, $\lambda^n - (\operatorname{tr}\boldsymbol{A})\lambda^{n-1} + \cdots + (-1)^n \det\boldsymbol{A} = 0$
- Proposition I: Let λ, μ be two distinct eigenvalues for $\boldsymbol{A} \in {}^n\mathbb{R}^n$, then $\mathscr{E}_\lambda \cap \mathscr{E}_\mu = \{\,\boldsymbol{0}\,\}$
  - Proof: Let $\boldsymbol{x} \in \mathscr{E}_\lambda \cap \mathscr{E}_\mu$, then $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\boldsymbol{A}\boldsymbol{x} = \mu\boldsymbol{x} \implies \lambda\boldsymbol{x} = \mu\boldsymbol{x} \implies (\lambda - \mu)\boldsymbol{x} = \boldsymbol{0}$, but $\lambda \neq \mu$ so $\boldsymbol{x} = \boldsymbol{0}$
- Note eigenvalues may be complex

Diagonalizability

- We have $\begin{cases} \boldsymbol{A}\boldsymbol{p}_1 = \lambda_1\boldsymbol{p}_1 \\ \vdots \\ \boldsymbol{A}\boldsymbol{p}_n = \lambda_n\boldsymbol{p}_n \end{cases} \implies \boldsymbol{A}\begin{bmatrix}\boldsymbol{p}_1 & \cdots & \boldsymbol{p}_n\end{bmatrix} = \begin{bmatrix}\boldsymbol{p}_1 & \cdots & \boldsymbol{p}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$
  - This can be written as $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{\Lambda}$
  - Suppose $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$ are linearly independent, then $\boldsymbol{P}$ is invertible, then $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$
    * This is called *diagonalization* since $\boldsymbol{\Lambda}$ is a diagonal matrix
- If we had $\dot{\boldsymbol{\eta}} = \boldsymbol{\Lambda}\boldsymbol{\eta} \implies \{\,\dot{\eta}_i = \lambda_i\eta_i\,\}$, which is a system of decoupled differential equations
- Definition: $\boldsymbol{P} \in {}^n\mathbb{R}^n$ *diagonalizes* $\boldsymbol{A} \in {}^n\mathbb{R}^n$ if $\boldsymbol{P}$ is invertible and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$
  - However, it's not always possible to find a set of $\boldsymbol{p}_n$ such that $\boldsymbol{P}$ is invertible, i.e. not all matrices are diagonalizable
- Theorem I: Diagonalization Theorem: The matrix $\boldsymbol{P} \in {}^n\mathbb{R}^n$ diagonalizes $\boldsymbol{A} \in {}^n\mathbb{R}^n$ (i.e. $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$) iff $\boldsymbol{P}$ the columns of $\boldsymbol{P}$ are eigenvectors of $\boldsymbol{A}$ that form a basis for ${}^n\mathbb{R}$
- Note that $P^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} \implies \boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$

Lecture 29, Mar 28, 2022

Diagonalization Properties

- Proposition II: Let $\boldsymbol{A}, \boldsymbol{T} \in {}^n\mathbb{R}^n$ and $\boldsymbol{T}$ be invertible, then $\boldsymbol{A}$ and $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ have the same characteristic polynomial and therefore same eigenvalues
  - $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ is known as a *similarity transformation* of $\boldsymbol{A}$
  - Proof:
$$\begin{aligned} &\det(\lambda\boldsymbol{1} - \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}) \\ &= \det(\lambda\boldsymbol{T}^{-1}\boldsymbol{T} - \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}) \\ &= \det(\boldsymbol{T}^{-1}(\lambda\boldsymbol{1} - \boldsymbol{A})\boldsymbol{T}) \\ &= \det(\boldsymbol{T}^{-1})\det(\lambda\boldsymbol{1} - \boldsymbol{A})\det(\boldsymbol{T}) \\ &= \det(\lambda\boldsymbol{1} - \boldsymbol{A}) \end{aligned}$$
- Theorem II: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ be diagonalizable; then:
  1. The characteristic equation for $\boldsymbol{A}$ can be written as $c_{\boldsymbol{A}}(\lambda) = \det(\lambda\boldsymbol{1} - \boldsymbol{A}) = \displaystyle\prod_{\alpha=1}^{n}(\lambda - \lambda_\alpha)$
     - If $\boldsymbol{A}$ is diagonalized by $\boldsymbol{P}$ then $c_{\boldsymbol{A}}(\lambda) = c_{\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}}(\lambda) = c_{\boldsymbol{\Lambda}}(\lambda)$
  2. $\det(\boldsymbol{A}) = \displaystyle\prod_{\alpha=1}^{n}\lambda_\alpha$

- Proof: $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} \implies \det(\boldsymbol{A}) = \det(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})$

$$= \det(\boldsymbol{P})\det(\boldsymbol{\Lambda})\det(\boldsymbol{P}^{-1})$$
$$= \det(\boldsymbol{\Lambda})$$
$$= \prod_{\alpha=1}^{n}(\lambda - \lambda_\alpha)$$

3. $\operatorname{tr}\boldsymbol{A} = \sum_{\alpha=1}^{n}\lambda_\alpha$
   - Proof: $\operatorname{tr}\boldsymbol{A} = \operatorname{tr}(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}) = \operatorname{tr}(\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{\Lambda}) = \operatorname{tr}(\boldsymbol{\Lambda}) = \sum_{\alpha=1}^{n}\lambda_\alpha$
   - Note $\operatorname{tr}(\boldsymbol{S}\boldsymbol{T}) = \operatorname{tr}(\boldsymbol{T}\boldsymbol{S})$

- Theorem II holds for all matrices, even ones that are not diagonalizable, we just currently cannot prove it
- It's important to note that repeated eigenvalues are counted multiple times

Lecture 30, Mar 29, 2022

Independence of Eigenspaces

- Theorem III: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ have r distinct eigenvalues ($r \leq n$) denoted $\lambda_1, \cdots, \lambda_r$, and let $\boldsymbol{x}_\alpha \in \mathscr{E}_{\lambda_\alpha}$ but $\boldsymbol{x}_\alpha \neq \boldsymbol{0}$; then $\{\, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_r \,\}$ is linearly independent
  - i.e. eigenvectors corresponding to different eigenvalues are always linearly independent
  - Proof by induction:
    * For $k = 1$, the set $\{\, \boldsymbol{x}_1 \,\}$ is linearly independent
    * Assume $\{\, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_k \,\}$ is linearly independent
    * Consider $\sum_{i=1}^{k+1}\mu_i\boldsymbol{x}_i = \boldsymbol{0}$

$$\implies \sum_{i=1}^{k+1}\mu_i\boldsymbol{A}\boldsymbol{x}_i = \boldsymbol{0}$$
$$\implies \sum_{i=1}^{k+1}\mu_i\lambda_i\boldsymbol{x}_i = \boldsymbol{0}$$
$$\implies \sum_{i=1}^{k+1}\mu_i\lambda_i\boldsymbol{x}_i - \lambda_{k+1}\sum_{i=1}^{k+1}\mu_i\boldsymbol{x}_i = \boldsymbol{0}$$
$$\implies \sum_{i=1}^{k+1}\mu_i(\lambda_i - \lambda_{k+1})\boldsymbol{x}_i = \boldsymbol{0}$$
$$\implies \sum_{i=1}^{k}\mu_i(\lambda_i - \lambda_{k+1})\boldsymbol{x}_i = \boldsymbol{0}$$
$$\implies \mu_1, \cdots, \mu_k = \boldsymbol{0}$$
$$\implies \mu_{k+1}\boldsymbol{x}_{k+1} = \boldsymbol{0}$$
$$\implies \mu_{k+1} = 0$$
$$\implies \{\, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{k+1} \,\} \text{ is linearly independent}$$

  - Corollary: If all the eigenvalues of $\boldsymbol{A}$ are distinct, then $\boldsymbol{A}$ is diagonalizable (since if $r = n$, we can pick a set of n independent eigenvectors, which must be a basis for ${}^n\mathbb{R}$)
    * However, if the eigenvalues are not distinct, that doesn't mean the matrix is not diagonalizable
- Lemma I: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ have r distinct eigenvalues and $\boldsymbol{x}_\alpha \in \mathscr{E}_{\lambda_\alpha}$, if $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_n = \boldsymbol{0}$ then $\boldsymbol{x}_\alpha = \boldsymbol{0}$
  - Proof:

    * Consider $\mu_1 \boldsymbol{x}_1 + \cdots + \mu_r \boldsymbol{x}_r = \mathbf{0}$
    * If $\boldsymbol{x}_\alpha \neq 0$ for some α then $\mu_\alpha = 0$ since $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r$ are independent
    * This contradicts $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_r = \mathbf{0}$
- Theorem IV: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ have r distinct eigenvalues and H_{λ_α} be a linearly independent set of eigenvectors from $\mathscr{E}_{\lambda_\alpha}$, then $H_{\lambda_1} \cup H_{\lambda_2} \cup \cdots \cup H_{\lambda_r}$ is linearly independent
  - Proof:
    * Let $H_{\lambda_\alpha} = \{ \boldsymbol{p}_{\alpha,1}, \cdots, \boldsymbol{p}_{\alpha,m_\alpha} \}$
    * $\sum_{j=1}^{m_1} \mu_{1,j}\boldsymbol{p}_{1,j} + \sum_{j=2}^{m_2} \mu_{2,j}\boldsymbol{p}_{2,j} + \cdots + \sum_{j=1}^{m_r} \mu_{r,j}\boldsymbol{p}_{r,j} = \mathbf{0}$
    * $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_r = 0 \implies \boldsymbol{x}_\alpha = \mathbf{0}$ by Lemma I
    * Since each sum adds up to zero, all μ are zero since each H is linearly independent
    * Therefore the union of all the sets is linearly independent

Lecture 31, Apr 1, 2022

Criteria for Diagonalizability

- $m_1 + m_2 + \cdots + m_r \leq n$ where $m_\alpha = \dim \mathscr{E}_{\lambda_\alpha}$ for $\boldsymbol{A} \in {}^n\mathbb{R}^n$
  - Proof: Let $\{ \boldsymbol{p}_{\alpha,j} \}$ be a set of eigenvectors in H_{λ_α}; $m_1 + \cdots + m_r$ is the total number of vectors in $H_{\lambda_1} \cup \cdots \cup H_{\lambda_\alpha}$; this cannot exceed n as that would violate the fundamental theorem
  - Corollary: If $m_1 + \cdots + m_r = n$, then $\boldsymbol{A}$ is diagonalizable
- Definition: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ have eigenvalues λ_α; then the *algebraic multiplicity* of λ_α is n_α, the highest power of $(\lambda - \lambda_\alpha)$ that divides the characteristic equation for $\boldsymbol{A}$
  - i.e. the algebraic multiplicity of λ_α is the number of times λ_α appears as a root of the characteristic polynomial
- Definition: The *geometric multiplicity* of λ_α is $m_\alpha = \dim \mathscr{E}_{\lambda_\alpha}$, i.e. the dimension of its eigenspace
  - Theorem VI: Diagonalization Test: $m_\alpha = n_\alpha$ for all α if and only if the matrix is diagonalizable (proven next lecture)
- Proposition III: Let $\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \boldsymbol{B} \\ \mathbf{0} & \boldsymbol{C} \end{bmatrix} \in {}^n\mathbb{R}^n$ and $\mathbf{1} \in {}^r\mathbb{R}^r$ (and $\boldsymbol{C} \in {}^{n-r}\mathbb{R}^{n-r}$), then $\det \boldsymbol{A} = \det \boldsymbol{C}$
  - Take $\boldsymbol{A}$ and reduce it into $\boldsymbol{A}' = \begin{bmatrix} \mathbf{1} & \boldsymbol{B}' \\ \mathbf{0} & \boldsymbol{C}' \end{bmatrix}$ where $\boldsymbol{C}$ is upper triangular
    * We can do this by some $\boldsymbol{E}$ without having to multiply any row by a scalar since we don't need the leading entries to be 1
  - $\det(\boldsymbol{A}') = \det(\boldsymbol{C}') = (-1)^p \det(\boldsymbol{C})$
  - $\det(\boldsymbol{A}') = (-1)^p \det(\boldsymbol{A})$ since the same operations were performed on $\boldsymbol{A}$
  - The minus signs cancel so $\det(\boldsymbol{A}) = \det(\boldsymbol{C})$
- Theorem V: Multiplicity Theorem: $1 \leq m_\alpha \leq n_\alpha$
  - Proof: Consider λ_α; Let $F = \{ \boldsymbol{f}_1, \cdots, \boldsymbol{f}_{m_\alpha} \}$ be a basis for $\mathscr{E}_{\lambda_\alpha}$ where $m_\alpha = \dim \mathscr{E}_{\lambda_\alpha}$
    * We can extend this basis for a basis for ${}^n\mathbb{R}$
    * Let $\boldsymbol{Q}$ be the transition matrix from F to $E_0 = \{ \boldsymbol{e}_1, \cdots, \boldsymbol{e}_{m_\alpha}, \cdots, \boldsymbol{e}_n \}$, the standard basis for ${}^n\mathbb{R}$, then $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{f}_1 & \cdots & \boldsymbol{f}_n \end{bmatrix}$ and $\boldsymbol{Q}\boldsymbol{e}_\alpha = \boldsymbol{f}_\alpha$, i.e. $\boldsymbol{Q}^{-1}\boldsymbol{f}_\alpha = \boldsymbol{e}_\alpha$
    * Consider $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{e}_{j_\alpha} = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{f}_{j_\alpha} = \lambda_\alpha \boldsymbol{Q}^{-1}\boldsymbol{f}_{j\alpha} = \lambda_\alpha \boldsymbol{e}_{j_\alpha}$ where $j_\alpha = 1, \cdots, m_\alpha$
    * So $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \begin{bmatrix} \lambda_\alpha \mathbf{1} & \boldsymbol{B} \\ \mathbf{0} & \boldsymbol{C} \end{bmatrix}$ where $\mathbf{1} \in {}^{m_\alpha}\mathbb{R}^{m_\alpha}$ since using the standard basis vectors we can pick out the λ_α for the first m_α columns
    * Consider
$$
\begin{aligned}
c_{\boldsymbol{A}}(\lambda) &= c_{\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}}(\lambda) \\
&= \det \begin{bmatrix} (\lambda - \lambda_\alpha)\mathbf{1} & -\boldsymbol{B} \\ \mathbf{0} & \lambda\mathbf{1} - \boldsymbol{C} \end{bmatrix} \\
&= (\lambda - \lambda_\alpha)^{m_\alpha} \det \begin{bmatrix} \mathbf{1} & -\boldsymbol{B} \\ \mathbf{0} & \lambda\mathbf{1} - \boldsymbol{C} \end{bmatrix} \\
&= (\lambda - \lambda_\alpha)^{m_\alpha} \det(\lambda\mathbf{1} - \boldsymbol{C})
\end{aligned}
$$

        - Note the first line relies on the similarity transformation preserving the characteristic equation
        - The last line relies on Prop. III
    * This shows us that we have to have at least m_α repeated roots of λ_α, so $m_\alpha \leq n_\alpha$
    * Since every eigenspace must have at least one nontrivial eigenvector $m_\alpha \geq 1$

Lecture 32, Apr 4, 2022

The Diagonalization Test

- Theorem VI: Diagonalization Test: Let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ with distinct eigenvalues $\lambda_1, \cdots, \lambda_r$, then $\boldsymbol{A}$ is diagonalizable if and only if $\forall\alpha, m_\alpha = n_\alpha$, where m_α is the geometric multiplicity and n_α is the algebraic multiplicity
  - Proof: [$\implies$] Let $\boldsymbol{A}$ be diagonalizable, then:
    * If $\boldsymbol{A}$ is diagonalizable then there are n linearly independent eigenvectors; let $E = E_{\lambda_1} \cup \cdots \cup E_{\lambda_r}$ be a linearly independent set of eigenvectors where E_{λ_α} is a basis for each eigenspace
    * Since E is a basis for ${}^n\mathbb{R}$, we have $n = |E|$ where E is the cardinality of E (i.e. number of elements)
    * Since $E_{\lambda_i} \cap E_{\lambda_j} = \varnothing$, so then $n = |E| = \displaystyle\sum_{\alpha=1}^{r} |E_{\lambda_\alpha}| = \sum_{\alpha=1}^{r} m_\alpha \leq \sum_{\alpha=1}^{r} n_\alpha = n$
        - Note $n_1 + n_2 + \cdots + n_r = n$
    * Therefore $\displaystyle\sum_{\alpha=1}^{r} m_\alpha = \sum_{\alpha=1}^{r} n_\alpha$, and since $m_\alpha \leq n_\alpha$ we must have $m_\alpha = n_\alpha$ for all α
  - Proof: [$\impliedby$] For $m_\alpha = n\alpha, \forall\alpha$:
    * $|E| = \displaystyle\sum_{\alpha=1}^{r} |E_{\lambda_\alpha}| = \sum_{\alpha=1}^{r} m_\alpha = \sum_{\alpha=1}^{r} n_\alpha = n$
    * Since $|E| = n$ there are n linearly independent eigenvectors, which span and form a basis for ${}^n\mathbb{R}$, so $\boldsymbol{A}$ is diagonalizable

Lecture 33, Apr 5, 2022

Solving Differential Equations with Diagonalization

- Given $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0$, how do we get $\boldsymbol{x}(t)$?
- Assume $\boldsymbol{A}$ is diagonalizable, then $\dot{\boldsymbol{x}} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\boldsymbol{x}$
- We can use $\boldsymbol{P}$ as a transition matrix; set $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{\eta}$, then
$$\begin{aligned} & \dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \\ \implies & \boldsymbol{P}\dot{\boldsymbol{\eta}} = (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})(\boldsymbol{P}\boldsymbol{\eta}) \\ \implies & \boldsymbol{P}\dot{\boldsymbol{\eta}} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{\eta} \\ \implies & \dot{\boldsymbol{\eta}} = \boldsymbol{\Lambda}\boldsymbol{\eta} \end{aligned}$$
  - Since $\boldsymbol{\Lambda}$ is diagonal, we have now decoupled the system!
  - Each equation becomes $\dot{\boldsymbol{\eta}}_\alpha = \lambda_\alpha \boldsymbol{\eta}_\alpha$, so each solution is $\boldsymbol{\eta}_\alpha(t) = c_\alpha e^{\lambda_\alpha t}$
- The full solution becomes $\boldsymbol{x}(t) = \boldsymbol{P}\boldsymbol{\eta}(t) = \begin{bmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_n \end{bmatrix} \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix} = c_1 \boldsymbol{p}_1 e^{\lambda_1 t} + \cdots + c_n \boldsymbol{p}_n e^{\lambda_n t}$
- Plugging the initial conditions $t = 0$ gives $c_1\boldsymbol{p}_1 + \cdots + c_n\boldsymbol{p}_n = \boldsymbol{x}_0 = \boldsymbol{P}\boldsymbol{c}$; solving the system gives the coefficients
- The eigenvalues λ are in the exponents, which dictate the speed at which the solution decays, or the frequency of oscillations
- The eigenvectors dictate the shape of the solution

Lecture 34, Apr 8, 2022

The Matrix Exponential

- Consider the scalar case where $\exp x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- Let $\boldsymbol{X} \in {}^n\mathbb{R}^n$, then define $e^{\boldsymbol{X}} = \exp \boldsymbol{X} = \sum_{k=0}^{\infty} \frac{\boldsymbol{X}^k}{k!}$
  - Note we have to define $\boldsymbol{X}^0 = \mathbf{1}$
- We know that in the scalar case $\dot{x} = ax$ has solution $x(t) = x_0 e^{at}$; can we do the same for the matrix exponential?
- $e^{\boldsymbol{A}t} = \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{k!} t^k \implies \frac{\mathrm{d}}{\mathrm{d}t} e^{\boldsymbol{A}t} = \sum_{k=1}^{\infty} \frac{\boldsymbol{A}^k}{(k-1)!} t^{k-1} = \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^{k+1}}{k!} t^k = \boldsymbol{A} \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{k!} t^k = \boldsymbol{A} e^{\boldsymbol{A}t}$
- Therefore in the general case of $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0$ the solution is $e^{\boldsymbol{A}t}\boldsymbol{x}_0$
- How do we actually compute $e^{\boldsymbol{A}t}$?
  - If $\boldsymbol{A}$ is diagonalizable, then $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$, so $\boldsymbol{A}^n = (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})(\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}) \cdots (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}) = \boldsymbol{P}\boldsymbol{\Lambda}^n\boldsymbol{P}^{-1}$
  - $\boldsymbol{\Lambda}^n$ is easy to compute, since $\boldsymbol{\Lambda}$ is diagonal, we simply take the diaongal entries to the nth power
- Using this result,

$$\begin{aligned}
e^{\boldsymbol{A}t} &= \sum_{k=0}^{\infty} \boldsymbol{P} \frac{\boldsymbol{\Lambda}^n}{k!} \boldsymbol{P}^{-1} t^k \\
&= \boldsymbol{P} \left(\sum_{k=0}^{\infty} \frac{\boldsymbol{\Lambda}^k}{k!} t^k \right) \boldsymbol{P}^{-1} \\
&= \boldsymbol{P} e^{\boldsymbol{\Lambda}t} \boldsymbol{P}^{-1} \\
&= \boldsymbol{P} \begin{bmatrix} \sum_{k=0}^{\infty} \frac{\lambda_2^k}{k!} t^k & 0 & \cdots \\ 0 & \sum_{k=0}^{\infty} \frac{\lambda_2^k}{k!} t^k & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \boldsymbol{P}^{-1} \\
&= \boldsymbol{P} \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots \\ 0 & e^{\lambda_2 t} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \boldsymbol{P}^{-1}
\end{aligned}$$

Lecture 35, Apr 11, 2022

Example Problems

- Consider the sequence: $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k, \boldsymbol{x}_k \in {}^n\mathbb{R}$, and let $\boldsymbol{A} \in {}^n\mathbb{R}^n$ be diagonalizable with real eigenvalues; show that if all $|\lambda_\alpha| < 1$, then $\boldsymbol{x}_k \to 0$ as $k \to \infty$
  - Note $\boldsymbol{x}_k = \boldsymbol{A}^k \boldsymbol{x}_0$
  - $\lim_{k\to\infty} \boldsymbol{x}_k = \lim_{k\to\infty} \boldsymbol{A}^k \boldsymbol{x}_0 = \lim_{k\to\infty} \boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1}\boldsymbol{x}_0$
  - $\lim_{k\to\infty} \boldsymbol{\Lambda}^k = \lim_{k\to\infty} \begin{bmatrix} \lambda_1^k & & \\ & \lambda_2^k & \\ & & \ddots \end{bmatrix} = \mathbf{0}$
  - Therefore $\lim_{k\to\infty} \boldsymbol{P}\boldsymbol{\Lambda}^k\boldsymbol{P}^{-1}\boldsymbol{x}_0 = \mathbf{0}$