Week 3

Matrices and linear transformations (Section 2.3)

Matrix as a "transformation:"

input	matrix	output
$\mathbf{x} \in \mathbb{R}^n$	$A \in \mathbb{R}^{m\times n}$	$A\mathbf{x} \in \mathbb{R}^m$

Definition 2.25: Let A be an $m \times n$ matrix. $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is the function defined by

$$T_A(\underbrace{\mathbf{x}}_{\in\mathbb{R}^n}) = \underbrace{A\mathbf{x}}_{\in\mathbb{R}^m}.$$

Example: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$$T_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \quad \text{("swap coordinates")}$$

Observation 2.26: Let A be an $m \times n$ matrix, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then

(i) $T_A(\mathbf{x} + \mathbf{y}) = T_A(\mathbf{x}) + T_A(\mathbf{y})$ and

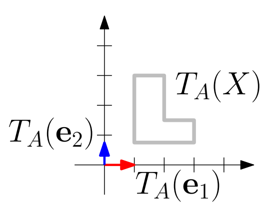
(ii) $T_A(\lambda\mathbf{x}) = \lambda T_A(\mathbf{x})$.

Combining (i) and (ii): $T_A(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda T_A(\mathbf{x}) + \mu T_A(\mathbf{y})$.

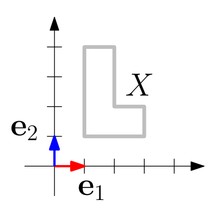
Proof. This just says (i) $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$ and (ii) $A(\lambda\mathbf{x}) = \lambda A\mathbf{x}$; both are true by the rules of vector addition, scalar multiplication, matrix multiplication. □

Transforming a set X of inputs: $T_A(X) := \{T_A(\mathbf{x}) : \mathbf{x} \in X\}$

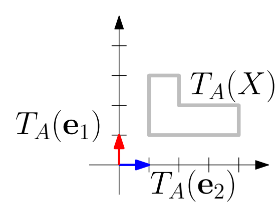
Examples $\mathbb{R}^2 \to \mathbb{R}^2$: $T_A(\mathbf{e}_1) = A\mathbf{e}_1, T_A(\mathbf{e}_2) = A\mathbf{e}_2$, the two columns of A



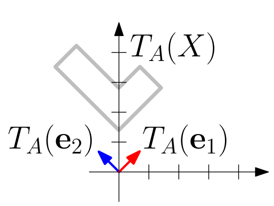
$A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{4} \end{bmatrix}$ $\longleftarrow$



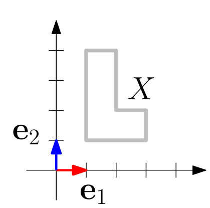
$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\longrightarrow$



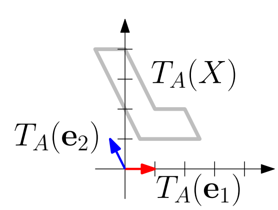
stretching — inputs — mirroring



$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ $\longleftarrow$

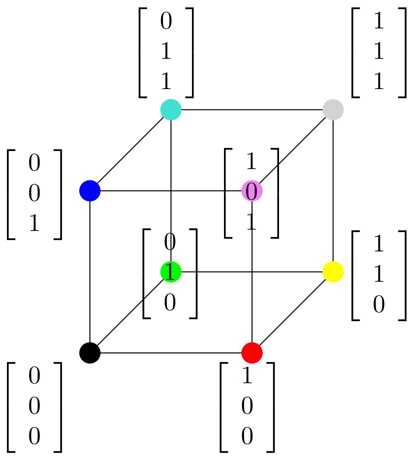


$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}$ $\longrightarrow$

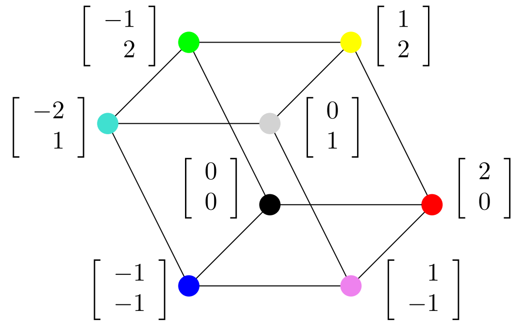


rotation — inputs — shear

Example $\mathbb{R}^3 \to \mathbb{R}^2$: *orthogonal projection*



$$A = \begin{bmatrix} 2 & -1 & -1 \\ 0 & 2 & -1 \end{bmatrix} \longrightarrow$$



inputs projection

Linear transformations:

Definition 2.27: A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $\lambda \in \mathbb{R}$,

(i) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and

(ii) $T(\lambda \mathbf{x}) = \lambda T(\mathbf{x})$.

Combining (i) and (ii): $T(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y})$.

All T_A's are linear transformations (Observation 2.26).

(i) and (ii): *axioms* of linear transformations

$$\begin{array}{ccccc} & \mathbf{x}, \mathbf{y} & \overset{T}{\longrightarrow} & T(\mathbf{x}), T(\mathbf{y}) & \\ + & \downarrow & & \downarrow & + \\ & \mathbf{x} + \mathbf{y} & \underset{T}{\longrightarrow} & T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) & \end{array} \qquad \begin{array}{ccccc} & \mathbf{x} & \overset{T}{\longrightarrow} & T(\mathbf{x}) & \\ \cdot\lambda & \downarrow & & \downarrow & \cdot\lambda \\ & \lambda\mathbf{x} & \underset{T}{\longrightarrow} & T(\lambda\mathbf{x}) = \lambda T(\mathbf{x}) & \end{array}$$

commutative diagrams: T "commutes" with $+$ and $\cdot$

Examples: $T = T_A$ for ...

- $T(\mathbf{x}) = \sum_{i=1}^n x_i$ — $A = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$
- $T(\mathbf{x}) = \mathbf{0}$ — $A = 0$
- $T(\mathbf{x}) = \mathbf{x}$ — $A = I$

Counterexamples: violation of ...

- $T(\mathbf{x}) = \sum_{i=1}^n |x_i|$ — (ii): if $\mathbf{x} \neq \mathbf{0}$ and $\lambda < 0$, then $T(\lambda \mathbf{x}) > 0$ but $\lambda T(\mathbf{x}) < 0$
- $T(\mathbf{x}) = \mathbf{u}$ with fixed $\mathbf{u} \neq \mathbf{0}$ — (i): $T(\mathbf{x} + \mathbf{y}) = \mathbf{u}$ but $T(\mathbf{x}) + T(\mathbf{y}) = 2\mathbf{u}$

Lemma 2.28: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\ell \in \mathbb{R}^n$ and $\lambda_1, \lambda_2, \ldots, \lambda_\ell \in \mathbb{R}$. Then

$$T\left(\sum_{j=1}^{\ell} \lambda_j \mathbf{x}_j\right) = \sum_{j=1}^{\ell} \lambda_j T(\mathbf{x}_j).$$

$\ell = 0$: $T(\mathbf{0}) = \mathbf{0}$, true by (ii) $\quad$ $\ell = 1$: (ii) $\quad$ $\ell = 2$: combination of (i) and (ii) $\quad$ $\ell > 2$?

Proof.

$$T\left(\sum_{j=1}^{\ell} \lambda_j \mathbf{x}_j\right) = T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j + \lambda_\ell \mathbf{x}_\ell\right) \overset{(i)}{=} T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j\right) + T\left(\lambda_\ell \mathbf{x}_\ell\right) \overset{(ii)}{=} T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j\right) + \lambda_\ell T\left(\mathbf{x}_\ell\right).$$

Same for $\ell - 1$:

$$T\left(\sum_{j=1}^{\ell} \lambda_j \mathbf{x}_j\right) = \underbrace{T\left(\sum_{j=1}^{\ell-2} \lambda_j \mathbf{x}_j\right) + \lambda_{\ell-1} T\left(\mathbf{x}_{\ell-1}\right)}_{T(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j)} + \lambda_\ell T\left(\mathbf{x}_\ell\right).$$

Repeating for $\ell - 2, \ldots, 1$:

$$T\left(\sum_{j=1}^{\ell} \lambda_j \mathbf{x}_j\right) = \underbrace{T\left(\sum_{j=1}^{0} \lambda_j \mathbf{x}_j\right)}_{T(\mathbf{0}) = \mathbf{0}} + \lambda_1 T(\mathbf{x}_1) + \ldots + \lambda_{\ell-1} T\left(\mathbf{x}_{\ell-1}\right) + \lambda_\ell T\left(\mathbf{x}_\ell\right) = \sum_{j=1}^{\ell} \lambda_j T(\mathbf{x}_j).$$

□

Proof by induction (without "repeating"):
For $\ell = 0$, prove it directly (*base case*): statement reads as $T(\mathbf{0}) = \mathbf{0}$ which is true.
For $\ell > 0$, prove the *induction step*: if the statement is true for $\ell - 1$ (*induction hypothesis*), then it is also true for ℓ.
Having done this, we know that the statement is true for all ℓ. Induction step:

$$\begin{aligned} T\left(\sum_{j=1}^{\ell} \lambda_j \mathbf{x}_j\right) &= T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j + \lambda_\ell \mathbf{x}_\ell\right) \overset{(i)}{=} T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j\right) + T\left(\lambda_\ell \mathbf{x}_\ell\right) \\ &\overset{(ii)}{=} T\left(\sum_{j=1}^{\ell-1} \lambda_j \mathbf{x}_j\right) + \lambda_\ell T\left(\mathbf{x}_\ell\right) \quad \text{(as before)} \\ &= \sum_{j=1}^{\ell-1} \lambda_j T(\mathbf{x}_j) \quad + \lambda_\ell T\left(\mathbf{x}_\ell\right) \quad \text{(induction hypothesis)} \\ &= \sum_{j=1}^{\ell} \lambda_j T(\mathbf{x}_j). \end{aligned}$$

□

The matrix of a linear transformation:

Theorem 2.29: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. There exists a unique $m \times n$ matrix A such that $T = T_A$.

Proof. For $T = T_A$, we need $T(\mathbf{e}_j) = T_A(\mathbf{e}_j) = A\mathbf{e}_j$ (j-th column of A), for all $j \in [n]$. Only candidate is

$$A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}.$$

This works:

$$\begin{aligned} T_A(\mathbf{x}) &= A\mathbf{x} && \text{(Definition 2.25, } T_A\text{)} \\ &= \sum_{j=1}^{n} x_j T(\mathbf{e}_j) && \text{(Definition 2.4, matrix-vector multiplication)} \\ &= T\left(\sum_{j=1}^{n} x_j \mathbf{e}_j\right) && \text{(Lemma 2.28)} \\ &= T(\mathbf{x}). \end{aligned}$$

□

Consequence: If we know $T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_n)$, we know $T(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Linear transformations and matrix multiplication:

Lemma 2.30: Let $T_A : \mathbb{R}^n \to \mathbb{R}^a$ and $T_B : \mathbb{R}^b \to \mathbb{R}^n$ be two linear transformations. Then

$$T_A(T_B(\mathbf{x})) = T_{AB}(\mathbf{x}).$$

Proof.

$$T_A(T_B(\mathbf{x})) = T_A(B\mathbf{x}) = A(B\mathbf{x}) = (AB)\mathbf{x} = T_{AB}(\mathbf{x}).$$

□

Can be used to prove generalized associativity, for example $(AB)(CD) = A((BC)D)$:

$$\begin{aligned} T_{(AB)(CD)}(\mathbf{x}) &= T_{AB}(T_{CD}(\mathbf{x})) = T_{AB}(T_C(T_D(\mathbf{x}))) = T_A(T_B(T_C(T_D(\mathbf{x})))) \\ T_{A((BC)D)}(\mathbf{x}) &= T_A(T_{(BC)D}(\mathbf{x})) = T_A(T_{BC}(T_D(\mathbf{x}))) = T_A(T_B(T_C(T_D(\mathbf{x})))) \end{aligned}$$

Same functions ⇒ same matrices (Theorem 2.29).

Systems of linear equations (Section 3.1)

$$\begin{aligned} D &= 2S \\ D &= C + 3 \\ D + S + C &= 17 \end{aligned}$$

children's age puzzle (Section 0.3)

$$\begin{aligned} x_1 - 2x_2 &= 0 \\ x_1 - x_3 &= 3 \\ x_1 + x_2 + x_3 &= 17 \end{aligned}$$

standard form ($x_1 = D, x_2 = S, x_3 = C$)

Definition 3.1: A *system of linear equations* in m equations and n variables $x_1, x_2, \ldots, x_n$ is of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

The a_{ij} and b_i: known real numbers The x_i: unknown real numbers (to be computed)
Matrix-vector form:

$$A\mathbf{x} = \mathbf{b} : \quad \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A,\ m\times n} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}\in\mathbb{R}^n} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}\in\mathbb{R}^m}.$$

A: *coefficient matrix* $\mathbf{x}$: *vector of variables* $\mathbf{b}$: *right-hand side*
Solving the system: compute $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$.

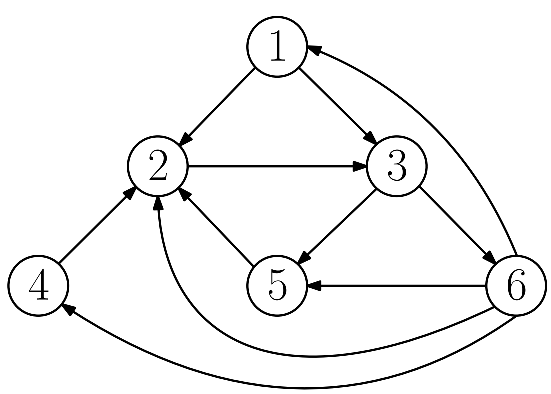
Children's age puzzle:

$$\underbrace{\begin{bmatrix} 1 & -2 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 0 \\ 3 \\ 17 \end{bmatrix}}_{\mathbf{b}}.$$

Observation 3.2: Let A be an $m \times n$ matrix. The columns of A are linearly independent if and only if the system $A\mathbf{x} = \mathbf{0}$ has a unique solution, $\mathbf{x} = \mathbf{0}$.

Proof. Unique solution $\Leftrightarrow$ $\mathbf{0}$ can only be written as a trivial linear combination of the colums $\Leftrightarrow$ columns are linearly independent (Lemma 1.19). □

The PageRank algorithm: works on *link graph* (circles: web pages, arrows: links)



Which page is most relevant?

Old school measure: number of *citations* (links to the page): page 2 (4 citations) wins.

PageRank principles:

- Citation from a relevant page counts more.
- Citation from a page that cites many pages counts less.

$\rightarrow$ relevance: sum of relevances of citing pages, divided by number of pages cited:

$$x_2 = \frac{x_1}{2} + x_4 + x_5 + \frac{x_6}{4} \quad \text{(same for the other 5 pages)}$$

System of 6 linear equations in 6 variables! But with useless solution $\mathbf{0}$.
Fix: use *damping factor* d close to 1 (for example, $d = 7/8$):

$$x_2 = (1-d) + d\left(\frac{x_1}{2} + x_4 + x_5 + \frac{x_6}{4}\right)$$

Unique solution (rounded): page 3 (rank 1.7307) wins.

$$x_1 = 0.31797, x_2 = 1.6761, x_3 = 1.7307, x_4 = 0.31797, x_5 = 1.0751, x_6 = 0.88217.$$

Computer vectors and matrices: how to store a system of linear equations?
$\mathbf{b} \in \mathbb{R}^m$: array $\texttt{b}$ with entries $\texttt{b}[0], \texttt{b}[1], \ldots, \texttt{b}[\texttt{m}-1]$.

$$\texttt{b} = \begin{bmatrix} \texttt{b}[0] \\ \texttt{b}[1] \\ \vdots \\ \texttt{b}[\texttt{m}-1] \end{bmatrix} \quad \text{(computer vector)}$$

Array indices start from 0, not from 1!
$A \in \mathbb{R}^{m\times n}$: array $\texttt{A}$ with m entries $\texttt{A}[0], \texttt{A}[1], \ldots, \texttt{A}[\texttt{m}-1]$. Each $\texttt{A}[\texttt{i}]$: array with n entries.

$$\texttt{A} = \begin{bmatrix} — & \texttt{A}[0] & — \\ — & \texttt{A}[1] & — \\ & \vdots & \\ — & \texttt{A}[\texttt{m}-1] & — \end{bmatrix} \quad \text{(computer matrix in row notation)}$$

$$\texttt{A}[\texttt{i}] = \begin{bmatrix} \texttt{A}[\texttt{i}][0] & \texttt{A}[\texttt{i}][1] & \cdots & \texttt{A}[\texttt{i}][\texttt{n}-1] \end{bmatrix} \quad \text{(computer row vector)}$$

$\texttt{A}[\texttt{i}][\texttt{j}]$ is the entry of $\texttt{A}$ in row i and column j (both counting from 0).