Matrix Calculations: Linear maps, bases, and matrix multiplication

A. Kissinger

Institute for Computing and Information Sciences
Radboud University Nijmegen

Version: spring 2017

Outline

From last time

- Vector spaces $V, W, \ldots$ are special kinds of sets whose elements are called *vectors*.
- Vectors can be added together, or multiplied by a real number, For $\boldsymbol{v}, \boldsymbol{w} \in V$, $a \in \mathbb{R}$:

$$\boldsymbol{v} + \boldsymbol{w} \in V \qquad\qquad a \cdot \boldsymbol{v} \in V$$

- The simplest examples are:

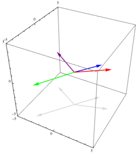
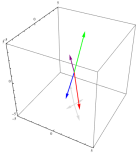
$$\mathbb{R}^n := \{(a_1, \ldots, a_n) \mid a_i \in \mathbb{R}\}$$

- Linear maps are special kinds of functions which satisfy two properties:

$$f(\boldsymbol{v} + \boldsymbol{w}) = f(\boldsymbol{v}) + f(\boldsymbol{w}) \qquad\qquad f(a \cdot \boldsymbol{v}) = a \cdot f(\boldsymbol{v})$$

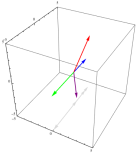
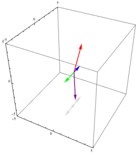
From last time

- Linear maps describe *transformations in space*, such as rotation:

 $\mapsto$ 

$$\mathrm{rx}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x \\ y\cos\theta - z\sin\theta \\ y\sin\theta + z\cos\theta \end{pmatrix}$$

- reflection and scaling:

 $\mapsto$ 

$$\mathrm{sy}\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x \\ (1/2)y \\ z \end{pmatrix}$$

Getting back to matrices

Q: So what is the relationship between this (cool) linear map stuff, and the (lets face it, kindof boring) stuff about matrices and linear equations from before?

A: Matrices are a convenient way to represent linear maps!

To get there, we need a new concept: *basis* of a vector space

Basis in space

- In $\mathbb{R}^3$ we can distinguish three special vectors:

$$(1,0,0) \in \mathbb{R}^3 \qquad (0,1,0) \in \mathbb{R}^3 \qquad (0,0,1) \in \mathbb{R}^3$$

- These vectors form a basis for $\mathbb{R}^3$, which means:
  1. These vectors *span* $\mathbb{R}^3$, which means each vector $(x, y, z) \in \mathbb{R}^3$ can be expressed as a linear combination of these three vectors:

$$\begin{aligned}(x, y, z) &= (x, 0, 0) + (0, y, 0) + (0, 0, z) \\ &= x \cdot (1,0,0) + y \cdot (0,1,0) + z \cdot (0,0,1)\end{aligned}$$

  2. Moreover, this set is as small as possible: no vectors are can be removed and still span $\mathbb{R}^3$.
- Note: condition (2) is equivalent to saying these vectors are linearly independent

Basis

Definition

Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ form a basis for a vector space V if these $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$

- are linearly independent, and
- span V in the sense that each $\boldsymbol{w} \in V$ can be written as linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, namely as:

$$\boldsymbol{w} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n \quad \text{for some} \quad a_1, \ldots, a_n \in \mathbb{R}$$

- These scalars a_i are uniquely determined by $\boldsymbol{w} \in V$ (see below)
- A space V may have several bases, but the number of elements of a basis for V is always the same; it is called the dimension of V, usually written as $\dim(V) \in \mathbb{N}$.

The standard basis for $\mathbb{R}^n$

- For the space $\mathbb{R}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}\}$ there is a standard choice of basis vectors:

  $$\boldsymbol{e}_1 := (1, 0, 0 \ldots, 0),\ \boldsymbol{e}_2 := (0, 1, 0, \ldots, 0),\ \cdots, \boldsymbol{e}_n := (0, \ldots, 0, 1)$$

- $\boldsymbol{e}_i$ has a 1 in the i-th position, and 0 everywhere else.
- We can easily check that these vectors are independent and span $\mathbb{R}^n$.
- This enables us to state precisely that $\mathbb{R}^n$ is n-dimensional.

An alternative basis for $\mathbb{R}^2$

- The standard basis for $\mathbb{R}^2$ is $(1, 0)$, $(0, 1)$.
- But many other choices are possible, eg. $(1, 1)$, $(1, -1)$
  - **independence**: if $a \cdot (1, 1) + b \cdot (1, -1) = (0, 0)$, then:

$$\begin{cases} a + b = 0 \\ a - b = 0 \end{cases} \quad \text{and thus} \quad \begin{cases} a = 0 \\ b = 0 \end{cases}$$

  - **spanning**: each point (x, y) can written in terms of $(1, 1), (1, -1)$, namely:

$$(x, y) = \tfrac{x+y}{2}(1, 1) + \tfrac{x-y}{2}(1, -1)$$

Uniqueness of representations

Theorem

- *Suppose V is a vector space, with basis $v_1, \ldots, v_n$*
- *assume $\boldsymbol{x} \in V$ can be represented in two ways:*

$$\boldsymbol{x} = a_1 \boldsymbol{v}_1 + \cdots + a_n \boldsymbol{v}_n \qquad \text{and also} \qquad x = b_1 \boldsymbol{v}_1 + \cdots + b_n \boldsymbol{v}_n$$

Then: $a_1 = b_1$ and . . . and $a_n = b_n$.

Proof: This follows from independence of $v_1, \ldots, v_n$ since:

$$\begin{aligned} \mathbf{0} \;=\; \boldsymbol{x} - \boldsymbol{x} \;&=\; \big(a_1 \boldsymbol{v}_1 + \cdots + a_n \boldsymbol{v}_n\big) - \big(b_1 \boldsymbol{v}_1 + \cdots + b_n \boldsymbol{v}_n\big) \\ &=\; (a_1 - b_1)\boldsymbol{v}_1 + \cdots + (a_n - b_n)\boldsymbol{v}_n. \end{aligned}$$

Hence $a_i - b_i = 0$, by independence, and thus $a_i = b_i$. ☺

Representing vectors

- Fixing a basis $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ therefore gives us a *unique* way to represent a vector $\boldsymbol{v} \in V$ as a list of numbers called *coordinates*:

$$\boldsymbol{v} = a_1 \boldsymbol{v}_1 + \cdots + a_n \boldsymbol{v}_n$$

New notation: $\boldsymbol{v} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}_{\mathcal{B}}$

- If $V = \mathbb{R}^n$, and $\mathcal{B}$ is the standard basis, this is just the vector itself:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

- ...but if $\mathcal{B}$ is not the standard basis, this can be different
- ...and if $V \neq \mathbb{R}^n$, a list of numbers is meaningless without fixing a basis.

What does it mean?

"The introduction of numbers as coordinates is an act of violence."

– Hermann Weyl

What does it mean?

- Space is (probably) real
- ...but coordinates (and hence bases) only exist in our head
- Choosing a basis amounts to fixing some directions in space we decide to call "**up**", "**right**", "**forward**", etc.
- Then a linear combination like:

$$\boldsymbol{v} = 5 \cdot \mathbf{up} + 3 \cdot \mathbf{right} - 2 \cdot \mathbf{forward}$$

  describes a point in space, mathematically.
- ...and it makes working with *linear maps* a *lot* easier

Linear maps and bases, example I

- Take the linear map $f((x_1, x_2, x_3)) = (x_1 - x_2,\ x_2 + x_3)$
- **Claim**: this map is entirely determined by what it does on the basis vectors $(1,0,0), (0,1,0), (0,0,1) \in \mathbb{R}^3$, namely:

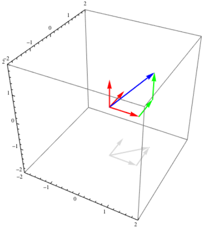
$$f((1,0,0)) = (1,0) \quad f((0,1,0)) = (-1,1) \quad f((0,0,1)) = (0,1).$$

- Indeed, using linearity:

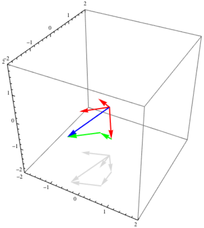
$$\begin{aligned} & f((x_1, x_2, x_3)) \\ &= f\Big((x_1,0,0) + (0,x_2,0) + (0,0,x_3)\Big) \\ &= f\Big(x_1 \cdot (1,0,0) + x_2 \cdot (0,1,0) + x_3 \cdot (0,0,1)\Big) \\ &= f\Big(x_1 \cdot (1,0,0)\Big) + f\Big(x_2 \cdot (0,1,0)\Big) + f\Big(x_3 \cdot (0,0,1)\Big) \\ &= x_1 \cdot f((1,0,0)) + x_2 \cdot f((0,1,0)) + x_3 \cdot f((0,0,1)) \\ &= x_1 \cdot (1,0) + x_2 \cdot (-1,1) + x_3 \cdot (0,1) \\ &= (x_1 - x_2,\ x_2 + x_3) \end{aligned}$$

Linear maps and bases, geometrically

"If we know how to transform *any* *set of axes for a space, we know how to transform everything."*



$\mapsto$



Linear maps and bases, example I (cntd)

- $f((x_1, x_2, x_3)) = (x_1 - x_2,\ x_2 + x_3)$ is totally determined by:

  $f((1,0,0)) = (1,0) \qquad f((0,1,0)) = (-1,1) \qquad f((0,0,1)) = (0,1)$

- We can organise this data into a 2×3 matrix:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

  The vector $f(\boldsymbol{v}_i)$, for basis vector $\boldsymbol{v}_i$, appears as the i-the column.

- Applying f can be done by a new kind of multiplication:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \cdot x_1 + -1 \cdot x_2 + 0 \cdot x_3 \\ 0 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ x_2 + x_3 \end{pmatrix}$$

Matrix-vector multiplication: Definition

Definition

For vectors $\boldsymbol{v} = (x_1, \ldots, x_n), \boldsymbol{w} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ define their inner product (or dot product) as the real number:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

Definition

If $\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$, then $\boldsymbol{w} := \boldsymbol{A} \cdot \boldsymbol{v}$

is the vector whose i-th element is the dot product of the i-th row of matrix $\boldsymbol{A}$ with the (input) vector $\boldsymbol{v}$.

Matrix-vector multiplication, explicitly

For $\boldsymbol{A}$ an $m \times n$ matrix, $\boldsymbol{B}$ a column vector of length n:

$$\boldsymbol{A} \cdot \boldsymbol{b} = \boldsymbol{c}$$

is a column vector of length m.

$$\begin{pmatrix} \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \vdots & \vdots \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \vdots \\ a_{i1}b_1 + \cdots + a_{in}b_n \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ c_i \\ \vdots \end{pmatrix}$$

$$c_i = \sum_{k=1}^{n} a_{ik} b_k$$

Representing linear maps

Theorem

For every linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, there exists an $m \times n$ matrix $\boldsymbol{A}$ where:

$$f(\boldsymbol{v}) = \boldsymbol{A} \cdot \boldsymbol{v}$$

(where "$\cdot$" is the matrix multiplication of $\boldsymbol{A}$ and a vector $\boldsymbol{v}$)

Proof. Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ be the standard basis for $\mathbb{R}^n$. $\boldsymbol{A}$ be the matrix whose i-th column is $f(\boldsymbol{e}_i)$. Then:

$$\boldsymbol{A} \cdot \boldsymbol{e}_i = \begin{pmatrix} a_{11}0 + \ldots + a_{1i}1 + \ldots + a_{1n}0 \\ \vdots \\ a_{m1}0 + \ldots + a_{mi}1 + \ldots + a_{mn}0 \end{pmatrix} = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} = f(\boldsymbol{e}_i)$$

Since it is enough to check basis vectors and $f(\boldsymbol{e}_i) = \boldsymbol{A} \cdot \boldsymbol{e}_i$, we are done. ☺

Matrix-vector multiplication, concretely

- Recall $f((x_1, x_2, x_3)) = (x_1 - x_2,\ x_2 + x_3)$ with matrix:

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

- We can directly calculate $f((1, 2, -1)) = (1 - 2, 2 - 1) = (-1, 1)$
- We can also get the same result by matrix-vector multiplication:

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

- This multiplication can be understood as: putting the argument values $x_1 = 1, x_2 = 2, x_3 = -1$ in variables of the underlying equations, and computing the outcome.

Another example, to learn the mechanics

$$\begin{pmatrix} 9 & 3 & 2 & 9 & 7 \\ 8 & 5 & 6 & 6 & 3 \\ 4 & 5 & 8 & 9 & 3 \\ 3 & 4 & 3 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ 5 \\ 2 \\ 5 \\ 7 \end{pmatrix}$$

$$= \begin{pmatrix} 9 \cdot 9 + 3 \cdot 5 + 2 \cdot 2 + 9 \cdot 5 + 7 \cdot 7 \\ 8 \cdot 9 + 5 \cdot 5 + 6 \cdot 2 + 6 \cdot 5 + 3 \cdot 7 \\ 4 \cdot 9 + 5 \cdot 5 + 8 \cdot 2 + 9 \cdot 5 + 3 \cdot 7 \\ 3 \cdot 9 + 4 \cdot 5 + 3 \cdot 2 + 3 \cdot 5 + 4 \cdot 7 \end{pmatrix}$$

$$= \begin{pmatrix} 81 + 15 + 4 + 45 + 49 \\ 72 + 25 + 12 + 30 + 21 \\ 36 + 25 + 16 + 45 + 21 \\ 27 + 20 + 6 + 15 + 28 \end{pmatrix} = \begin{pmatrix} 194 \\ 160 \\ 143 \\ 96 \end{pmatrix}$$

Linear map from matrix

- We have seen how a linear map can be described via a matrix
- One can also read each matrix as a linear map

Example

- Consider the matrix $\begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & -3 \end{pmatrix}$
- It has 3 columns/inputs and two rows/outputs. Hence it describes a map $f\colon \mathbb{R}^3 \to \mathbb{R}^2$
- Namely: $f((x_1, x_2, x_3)) = (2x_1 - x_3,\ 5x_1 + x_2 - 3x_3)$.

Examples of linear maps and matrices I

Projections are linear maps that send higher-dimensional vectors to lower ones. Consider $f : \mathbb{R}^3 \to \mathbb{R}^2$

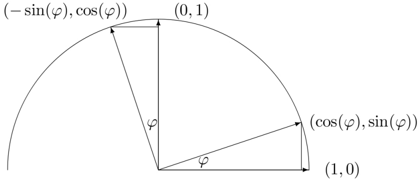
$$f(\begin{pmatrix} x \\ y \\ z \end{pmatrix}) = \begin{pmatrix} x \\ y \end{pmatrix}.$$

f maps $3d$ space to the the $2d$ plane.
The matrix of f is the following 2×3 matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Examples of linear maps and matrices II

We have already seen: Rotation over an angle φ is a linear map



This rotation is described by $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$f((x,y)) = \big(\, x\cos(\varphi) - y\sin(\varphi),\; x\sin(\varphi) + y\cos(\varphi)\,\big)$$

The matrix that describes f is

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}.$$

Example: systems of equations

$$\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & b_1 \\ & \vdots & \\ a_{m1}x_1 + \cdots + a_{mn}x_n & = & b_m \end{array} \quad \Rightarrow \quad \begin{array}{c} \boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b} \\ \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \vdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \end{array}$$

$$\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & 0 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n & = & 0 \end{array} \quad \Rightarrow \quad \begin{array}{c} \boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{0} \\ \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \end{array}$$

General vector spaces

- We can also represent linear maps $f : V \to W$ between general vector spaces (not just $\mathbb{R}^n$)
- But we must fix bases for both spaces:

$$\mathcal{B} := \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$$
$$\mathcal{C} := \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\} \subset W$$

- Then:

$$f(\boldsymbol{x}) = \boldsymbol{A} \cdot \boldsymbol{x}$$

where $\boldsymbol{A}$ is the matrix whose i-th column is $f(\boldsymbol{v}_i)$, written in terms of basis $\mathcal{C}$:

$$f(\boldsymbol{v}_i) = a_{1i}\boldsymbol{w}_1 + \ldots + a_{mi}\boldsymbol{w}_m = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}_{\mathcal{C}}$$

Matrix summary

- Fix bases $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$ and $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\} \subset W$
- Every linear map $f \colon V \to W$ can be represented by a matrix, and every matrix represents a linear map:

$$f(\boldsymbol{v}) = \boldsymbol{A} \cdot \boldsymbol{v}$$

- The i-th column of $\boldsymbol{A}$ is $f(\boldsymbol{v}_i)$, wrt. the basis $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m$ of W
- This matrix of f depends on the choice of basis: for different bases of V and W a different matrix is obtained
- For $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, we often use the standard basis, in which case the i-th column of $\boldsymbol{A}$ is just $f(\boldsymbol{e}_i)$.

Matrix multiplication

- Consider linear maps g, f represented by matrices $\boldsymbol{A}$, $\boldsymbol{B}$:

$$g(\boldsymbol{v}) = \boldsymbol{A} \cdot \boldsymbol{v} \qquad\qquad f(\boldsymbol{w}) = \boldsymbol{B} \cdot \boldsymbol{w}$$

- Can we find a matrix $\boldsymbol{C}$ that represents their composition?

$$g(f(\boldsymbol{v})) = \boldsymbol{C} \cdot \boldsymbol{v}$$

- Let's try:

$$g(f(\boldsymbol{v})) = g(\boldsymbol{B} \cdot \boldsymbol{v}) = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{v}) \overset{(*)}{=} (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{v}$$

(where step $(*)$ is currently 'wishful thinking')

- Great! Let $\boldsymbol{C} := \boldsymbol{A} \cdot \boldsymbol{B}$.
- But we don't know what "·" means for two matrices yet...

Matrix multiplication

- Solution: generalise from $\boldsymbol{A} \cdot \boldsymbol{v}$
- A vector is a matrix with one column:

  The number in the i-th row and the first column of $\boldsymbol{A} \cdot \boldsymbol{v}$ is the dot product of the i-th row of $\boldsymbol{A}$ with the first column of $\boldsymbol{v}$.
- So for matrices $\boldsymbol{A}, \boldsymbol{B}$:

  The number in the i-th row and the j-th column of $\boldsymbol{A} \cdot \boldsymbol{B}$ is the dot product of the i-th row of $\boldsymbol{A}$ with the j-th column of $\boldsymbol{B}$.

Matrix multiplication

For $\boldsymbol{A}$ an $m \times n$ matrix, $\boldsymbol{B}$ an $n \times p$ matrix:

$$\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{C}$$

is an $m \times p$ matrix.

$$\begin{pmatrix} \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \vdots & \vdots \end{pmatrix} \cdot \begin{pmatrix} \cdots & b_{j1} & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & b_{jn} & \cdots \end{pmatrix} = \begin{pmatrix} \ddots & \vdots & \iddots \\ \cdots & c_{ij} & \cdots \\ \iddots & \vdots & \ddots \end{pmatrix}$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Special case: vectors

For $\boldsymbol{A}$ an $m \times n$ matrix, $\boldsymbol{B}$ an $n \times 1$ matrix:

$$\boldsymbol{A} \cdot \boldsymbol{b} = \boldsymbol{c}$$

is an $m \times 1$ matrix.

$$\begin{pmatrix} \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \vdots & \vdots \end{pmatrix} \cdot \begin{pmatrix} b_{11} \\ \vdots \\ b_{n1} \end{pmatrix} = \begin{pmatrix} \vdots \\ c_{i1} \\ \vdots \end{pmatrix}$$

$$c_{i1} = \sum_{k=1}^{n} a_{ik} b_{k1}$$

Matrix composition

Theorem

Matrix composition is associative:

$$(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$$

Proof. Let $\boldsymbol{X} := (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C}$. This is a matrix with entries:

$$x_{ip} = \sum_k a_{ik} b_{kp}$$

Then, the matrix entries of $\boldsymbol{X} \cdot \boldsymbol{C}$ are:

$$\sum_p x_{ip} c_{pj} = \sum_p \left(\sum_k a_{ik} b_{kp} \right) c_{pk} = \sum_{kp} a_{ik} b_{kp} c_{pk}$$

(because sums can always be pulled outside, and combined)

Associativity of matrix composition

Proof (cont'd). Now, let $\boldsymbol{Y} := \boldsymbol{B} \cdot \boldsymbol{C}$. This has matrix entries:

$$y_{kj} = \sum_p b_{kp} c_{pj}$$

Then, the matrix entries of $\boldsymbol{A} \cdot \boldsymbol{Y}$ are:

$$\sum_k a_{ik} y_{kj} = \sum_k a_{ik} \left(\sum_p b_{kp} c_{pj} \right) = \sum_{kp} a_{ik} b_{kp} c_{pk}$$

...which is the same as before! So:

$$(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{X} \cdot \boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{Y} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$$

☺

So we can drop those pesky parentheses:

$$\boldsymbol{A} \cdot \boldsymbol{B} \cdot \boldsymbol{C} := (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$$

Matrix product and composition

Corollary

The composition of linear maps is given by matrix product.

Proof. Let $g(\boldsymbol{w}) = \boldsymbol{A} \cdot \boldsymbol{w}$ and $f(\boldsymbol{v}) = \boldsymbol{B} \cdot \boldsymbol{v}$. Then:

$$g(f(\boldsymbol{v})) = g(\boldsymbol{B} \cdot \boldsymbol{v}) = \boldsymbol{A} \cdot \boldsymbol{B} \cdot \boldsymbol{v}$$

☺

No wishful thinking necessary!

Example 1

Consider the following two linear maps, and their associated matrices:

$$\mathbb{R}^3 \xrightarrow{f} \mathbb{R}^2 \qquad \mathbb{R}^2 \xrightarrow{g} \mathbb{R}^2$$

$$f((x_1, x_2, x_3)) = (x_1 - x_2,\ x_2 + x_3) \quad g((y_1, y_2)) = (2y_1 - y_2, 3y_2)$$

$$\boldsymbol{M}_f = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{M}_g = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix}$$

We can compute the composition directly:

$$\begin{aligned} (g \circ f)((x_1, x_2, x_3)) &= g\big(f((x_1, x_2, x_3))\big) \\ &= g((x_1 - x_2,\ x_2 + x_3)) \\ &= (\,2(x_1 - x_2) - (x_2 + x_3),\ 3(x_2 + x_3)\,) \\ &= (\,2x_1 - 3x_2 - x_3,\ 3x_2 + 3x_3\,) \end{aligned}$$

So:

$$\boldsymbol{M}_{g \circ f} = \begin{pmatrix} 2 & -3 & -1 \\ 0 & 3 & 3 \end{pmatrix}$$

...which is just the product of the matrices: $\boldsymbol{M}_{g \circ f} = \boldsymbol{M}_g \cdot \boldsymbol{M}_f$

Note: matrix composition is not commutative

In general, $\boldsymbol{A} \cdot \boldsymbol{B} \neq \boldsymbol{B} \cdot \boldsymbol{A}$

For instance: Take $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then:

$$\begin{aligned} \boldsymbol{A} \cdot \boldsymbol{B} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 0 + 0 \cdot -1 & 1 \cdot 1 + 0 \cdot 0 \\ 0 \cdot 0 + -1 \cdot -1 & 0 \cdot 1 + -1 \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{B} \cdot \boldsymbol{A} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 0 + 1 \cdot -1 \\ -1 \cdot 1 + 0 \cdot 0 & -1 \cdot 0 + 0 \cdot -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \end{aligned}$$

But it is...

...associative, as we've already seen:

$$\boldsymbol{A} \cdot \boldsymbol{B} \cdot \boldsymbol{C} := (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$$

It also has a *unit* given by the *identity matrix* $\boldsymbol{I}$:

$$\boldsymbol{A} \cdot \boldsymbol{I} = \boldsymbol{I} \cdot \boldsymbol{A} = \boldsymbol{A}$$

where:

$$\boldsymbol{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Example: political swingers, part I

- We take an extremely crude view on politics and distinguish only left and right wing political supporters
- We study changes in political views, per year
- Suppose we observe, for each year:
  - 80% of lefties remain lefties and 20% become righties
  - 90% of righties remain righties, and 10% become lefties

Questions . . .

- start with a population $L = 100, R = 150$, and compute the number of lefties and righties after one year;
- similarly, after 2 years, and 3 years, . . .
- Find a convenient way to represent these computations.

Political swingers, part II

- So if we start with a population $L = 100, R = 150$, then after one year we have:
  - lefties: $0.8 \cdot 100 + 0.1 \cdot 150 = 80 + 15 = 95$
  - righties: $0.2 \cdot 100 + 0.9 \cdot 150 = 20 + 135 = 155$
- Two observations:
  - this looks like a matrix-vector multiplication
  - long-term developments can be calculated via iterated matrices

Political swingers, part III

- We can write the political transition matrix as

$$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix}$$

- If $\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} 100 \\ 150 \end{pmatrix}$, then after one year we have:

$$\begin{aligned} \boldsymbol{P} \cdot \begin{pmatrix} 100 \\ 150 \end{pmatrix} &= \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 100 \\ 150 \end{pmatrix} \\ &= \begin{pmatrix} 0.8 \cdot 100 + 0.1 \cdot 150 \\ 0.2 \cdot 100 + 0.9 \cdot 150 \end{pmatrix} = \begin{pmatrix} 95 \\ 155 \end{pmatrix} \end{aligned}$$

- After two years we have:

$$\begin{aligned} \boldsymbol{P} \cdot \begin{pmatrix} 95 \\ 155 \end{pmatrix} &= \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 95 \\ 155 \end{pmatrix} \\ &= \begin{pmatrix} 0.8 \cdot 95 + 0.1 \cdot 155 \\ 0.2 \cdot 95 + 0.9 \cdot 155 \end{pmatrix} = \begin{pmatrix} 91.5 \\ 158.5 \end{pmatrix} \end{aligned}$$

Political swingers, part IV

The situation after two years is obtained as:

$$\boldsymbol{P} \cdot \boldsymbol{P} \cdot \begin{pmatrix} L \\ R \end{pmatrix} = \underbrace{\begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix}}_{\text{do this multiplication first}} \cdot \begin{pmatrix} L \\ R \end{pmatrix}$$

$$= \begin{pmatrix} 0.8 \cdot 0.8 + 0.1 \cdot 0.2 & 0.8 \cdot 0.1 + 0.1 \cdot 0.9 \\ 0.2 \cdot 0.8 + 0.9 \cdot 0.2 & 0.2 \cdot 0.1 + 0.9 \cdot 0.9 \end{pmatrix} \cdot \begin{pmatrix} L \\ R \end{pmatrix}$$

$$= \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix} \cdot \begin{pmatrix} L \\ R \end{pmatrix}$$

The situation after n years is described by the n-fold iterated matrix:

$$\boldsymbol{P}^n = \underbrace{\boldsymbol{P} \cdot \boldsymbol{P} \cdots \boldsymbol{P}}_{n \text{ times}}$$

Political swingers, part V

Interpret the following iterations:

$$P^2 = P \cdot P = \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix}$$

$$P^3 = P \cdot P \cdot P = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix}$$

$$= \begin{pmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{pmatrix}$$

$$P^4 = P \cdot P \cdot P \cdot P = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{pmatrix}$$

$$= \begin{pmatrix} 0.4934 & 0.2533 \\ 0.5066 & 0.7467 \end{pmatrix}$$

Etc. It looks like P^{100} is going to be hard to calculate. Is there an easier way to do this? (**Spoiler alert:** Yes! But you'll have to wait 2 weeks...)