Matrix Calculations: Kernels & Images, Matrix Multiplication

A. Kissinger (and H. Geuvers)

Institute for Computing and Information Sciences – Intelligent Systems
Radboud University Nijmegen

Version: spring 2016

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

- Whereas there exist LOTS of functions between the sets V and W...
- ...there actually aren't that many 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 $\boldsymbol{v}$)

- More generally, every linear map $f : V \to W$ is representable as a matrix, but you have to fix a basis for V and W first:

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \in V \qquad \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\} \in W$$

- ...whereas in $\mathbb{R}^n$ there is an obvious choice:

$$\{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, \ldots, 0, 1)\} \in \mathbb{R}^n$$

Matrix-vector multiplication

For a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{v}$, $\boldsymbol{w} := \boldsymbol{A} \cdot \boldsymbol{v}$ is the vector whose i-th row is the dot product of the i-th row of $\boldsymbol{A}$ with $\boldsymbol{v}$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + \ldots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + \ldots + a_{mn}v_n \end{pmatrix}$$

i.e. $w_i \;:=\; a_{11}v_1 + \ldots + a_{1n}v_n \;=\; \sum_{j=1}^{n} a_{ij}v_j.$

Example: systems of equations

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

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

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 "$\cdot$" 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:

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

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

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

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:

$$\begin{aligned}
\boldsymbol{P} \cdot \boldsymbol{P} \cdot \begin{pmatrix} L \\ R \end{pmatrix} &= \underbrace{\begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix}}_{\text{do this multiplication first}} \cdot \begin{pmatrix} L \\ R \end{pmatrix} \\
&= \begin{pmatrix} 0.8 \cdot 0.8 + 0.1 \cdot 0.2 & 0.8 \cdot 0.1 + 0.1 \cdot 0.9 \\ 0.2 \cdot 0.8 + 0.9 \cdot 0.2 & 0.2 \cdot 0.1 + 0.9 \cdot 0.9 \end{pmatrix} \cdot \begin{pmatrix} L \\ R \end{pmatrix} \\
&= \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix} \cdot \begin{pmatrix} L \\ R \end{pmatrix}
\end{aligned}$$

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:

$$
\begin{aligned}
\boldsymbol{P}^2 &= \boldsymbol{P}\cdot\boldsymbol{P} = \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix} \\
\boldsymbol{P}^3 &= \boldsymbol{P}\cdot\boldsymbol{P}\cdot\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.66 & 0.17 \\ 0.34 & 0.83 \end{pmatrix} \\
&= \begin{pmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{pmatrix} \\
\boldsymbol{P}^4 &= \boldsymbol{P}\cdot\boldsymbol{P}\cdot\boldsymbol{P}\cdot\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.562 & 0.219 \\ 0.438 & 0.781 \end{pmatrix} \\
&= \begin{pmatrix} 0.4934 & 0.2533 \\ 0.5066 & 0.7467 \end{pmatrix}
\end{aligned}
$$

Etc. Does this stabilise? We'll talk about *fixed points* later on...

Solving equations the old fashioned way...

- We now know that systems of equations look like this:

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

- The goal is to solve for $\boldsymbol{x}$, in terms of $\boldsymbol{A}$ and $\boldsymbol{b}$.
- Here comes some more wishful thinking:

$$\boldsymbol{x} = \frac{1}{\boldsymbol{A}} \cdot \boldsymbol{b}$$

- Well, we can't really *divide* by a matrix, but if we are lucky, we can find another matrix called $\boldsymbol{A}^{-1}$ which acts like $\frac{1}{\boldsymbol{A}}$.

Inverse

Definition

The *inverse* of a matrix $\boldsymbol{A}$ is another matrix $\boldsymbol{A}^{-1}$ such that:

$$\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{I}$$

- Not all matrices have inverses, but when they do, we are happy, because:

$$\begin{aligned} \boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b} \quad &\Longrightarrow \quad \boldsymbol{A}^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{A}^{-1} \cdot \boldsymbol{b} \\ &\Longrightarrow \quad \boldsymbol{x} = \boldsymbol{A}^{-1} \cdot \boldsymbol{b} \end{aligned}$$

- So, how do we compute the inverse of a matrix?

Remember me?

Gaussian elimination as matrix multiplication

- Each step of Gaussian elimination can be represented by a matrix multiplication:

$$\boldsymbol{A} \Rightarrow \boldsymbol{A}' \qquad\qquad \boldsymbol{A}' := \boldsymbol{G} \cdot \boldsymbol{A}$$

- For instance, multiplying the i-th row by c is given by:

$$\boldsymbol{G}_{(R_i := cR_i)} \cdot \boldsymbol{A}$$

  where $\boldsymbol{G}_{(R_i := cR_i)}$ is just like the identity matrix, but $g_{ii} = c$.
- Exercise. What are the other Gaussian elimination matrices?

$$\boldsymbol{G}_{(R_i \leftrightarrow R_j)} \qquad\qquad \boldsymbol{G}_{(R_i := R_i + cR_j)}$$

Reduction to Echelon form

- The idea: treat $\boldsymbol{A}$ as a coefficient matrix, and compute its reduced Echelon form
- If the Echelon form of $\boldsymbol{A}$ has n pivots, then its reduced Echelon form is the identity matrix:

$$\boldsymbol{A} \Rightarrow \boldsymbol{A}_1 \Rightarrow \boldsymbol{A}_2 \Rightarrow \cdots \Rightarrow \boldsymbol{A}_p = \boldsymbol{I}$$

- Now, we can use our Gauss matrices to remember what we did:

$$\begin{aligned}
\boldsymbol{A}_1 &:= \boldsymbol{G}_1 \cdot \boldsymbol{A} \\
\boldsymbol{A}_2 &:= \boldsymbol{G}_2 \cdot \boldsymbol{G}_1 \cdot \boldsymbol{A} \\
&\cdots \\
\boldsymbol{A}_p &:= \boldsymbol{G}_p \cdots \boldsymbol{G}_1 \cdot \boldsymbol{A} = \boldsymbol{I}
\end{aligned}$$

Computing the inverse

- A ha!

$$\mathbf{G}_p \cdots \mathbf{G}_1 \cdot \mathbf{A} = \mathbf{I} \qquad \Longrightarrow \qquad \mathbf{A}^{-1} = \mathbf{G}_p \cdots \mathbf{G}_1$$

- So all we have to do is construct p different matrices and multiply them all together!
- Since I already have plans for this afternoon, lets take a shortcut:

Theorem

For $\mathbf{C}$ a matrix and $(\mathbf{A}|\mathbf{B})$ an augmented matrix:

$$\mathbf{C} \cdot (\mathbf{A}|\mathbf{B}) = (\mathbf{C} \cdot \mathbf{A} \mid \mathbf{C} \cdot \mathbf{B})$$

Computing the inverse

- Since Gaussian elimination is just multiplying by a certain matrix on the left...

$$\boldsymbol{A} \Rightarrow \boldsymbol{G} \cdot \boldsymbol{A}$$

- ...doing Gaussian elimination (for $\boldsymbol{A}$) on an augmented matrix applies $\boldsymbol{G}$ to both parts:

$$(\boldsymbol{A}|\boldsymbol{B}) \Rightarrow (\boldsymbol{G} \cdot \boldsymbol{A} \mid \boldsymbol{G} \cdot \boldsymbol{B})$$

- So, if $\boldsymbol{G} = \boldsymbol{A}^{-1}$:

$$(\boldsymbol{A}|\boldsymbol{B}) \Rightarrow (\boldsymbol{A}^{-1} \cdot \boldsymbol{A} \mid \boldsymbol{A}^{-1} \cdot \boldsymbol{B}) = (\boldsymbol{I}| \; \boldsymbol{A}^{-1} \cdot \boldsymbol{B})$$

Computing the inverse

- We already (secretly) used this trick to solve:

$$\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b} \qquad \Longrightarrow \qquad \boldsymbol{x} = \boldsymbol{A}^{-1} \cdot \boldsymbol{b}$$

- Here, we are only interested in the vector $\boldsymbol{A}^{-1} \cdot \boldsymbol{b}$
- Which is exactly what Gaussian elimination on the augmented matrix gives us:

$$(\boldsymbol{A}|\boldsymbol{b}) \Rightarrow (\boldsymbol{I}|\ \boldsymbol{A}^{-1} \cdot \boldsymbol{b})$$

- To get the entire matrix, we just need to choose something clever to the right of the line
- Like this:

$$(\boldsymbol{A}|\boldsymbol{I}) \Rightarrow (\boldsymbol{I}|\ \boldsymbol{A}^{-1} \cdot \boldsymbol{I}) = (\boldsymbol{I}|\ \boldsymbol{A}^{-1})$$

Computing the inverse: example

For example, we compute the inverse of:

$$\boldsymbol{A} := \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

as follows:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right)$$

So:

$$\boldsymbol{A}^{-1} := \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Computing the inverse: non-example

Unlike transpose, not every matrix has an inverse.
For example, if we try to compute the inverse for:

$$\boldsymbol{B} := \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

we have:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{array}\right)$$

We don't have enough pivots to continue reducing. So $\boldsymbol{B}$ does not have an inverse.

Subspace definition

Definition

A subset $S \subseteq V$ of a vector space V is called a (linear) subspace if S is closed under addition and scalar multiplication:

- $\mathbf{0} \in S$
- $\boldsymbol{v}, \boldsymbol{v}' \in S$ implies $\boldsymbol{v} + \boldsymbol{v}' \in S$
- $\boldsymbol{v} \in S$ and $a \in \mathbb{R}$ implies $a \cdot \boldsymbol{v} \in S$.

Note

- A subspace $S \subseteq V$ is a vector space itself, and thus also has a basis.
- Also S has its own dimension, where $\dim(S) \leq \dim(V)$.

Subspace examples

1. Earlier we saw that the subset of solutions of a system of equations is closed under addition and (scalar) multiplication, and thus is a linear subspace.

2. The diagonal $D = \{(x, x) \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$ is a linear subspace:
    - if $(x_1, x_1), (x_2, x_2) \in D$, then also $(x_1, x_1) + (x_2, x_2) = (x_1 + x_2, x_1 + x_2) \in D$
    - if $(x, x) \in D$ and $a \in \mathbb{R}$, also $a \cdot (x, x) = (a \cdot x, a \cdot x) \in D$

Also:

- D has a single vector as basis, for example $(1, 1)$
- thus, D has dimension 1

Basis for subspaces

Let the space V have dimension n, and a subspace $S \subseteq V$ dimension p, where $p \leq n$. Then:

- any set of $> p$ vectors in S is linearly dependent
- any set of $< p$ vectors in S does not span S
- any set of p independent vectors in S is a basis for S
- any set of p vectors that spans S is a basis for S

Kernel and image: definitions

Definition

Let $f\colon V \to W$ be a linear map

- the kernel of f is the subset of V given by:

$$\ker(f) = \{\boldsymbol{v} \in V \mid f(\boldsymbol{v}) = \boldsymbol{0}\}$$

- the image of f is the subset of W given by:

$$\mathrm{im}(f) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in V\}$$

Example

Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = (x,0)$

- the kernel is $\{(x,y) \in \mathbb{R}^2 \mid f(x,y) = (0,0)\}$, which is $\{(0,y) \mid y \in \mathbb{R}\}$, i.e. the y-axis.
- the image is the x-axis $\{(x,0) \mid x \in \mathbb{R}\}$

Kernels and images are subspaces

Theorem

For a linear map $f\colon V \to W$,

- $\ker(f) = \{\boldsymbol{v} \mid f(\boldsymbol{v}) = 0\} \subseteq V$ *is a linear subspace*
- $\mathrm{im}(f) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in V\} \subseteq W$ *is a linear subspace.*

Proof: We check two cases (do the others yourself!)

- **Closure of** $\ker(f)$ **under addition**: if $\boldsymbol{v}, \boldsymbol{v}' \in \ker(f)$, then $f(\boldsymbol{v}) = \mathbf{0}$ and $f(\boldsymbol{v}') = \mathbf{0}$. By linearity of f, $f(\boldsymbol{v} + \boldsymbol{v}') = f(\boldsymbol{v}) + f(\boldsymbol{v}') = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $\boldsymbol{v} + \boldsymbol{v}' \in \ker(f)$.
- **Closure of** $\mathrm{im}(f)$ **under scalar multiplication**: Assume $\boldsymbol{w} \in \mathrm{im}(f)$, say $\boldsymbol{w} = f(\boldsymbol{v})$, and $a \in \mathbb{R}$. Again by linearity: $a \cdot \boldsymbol{w} = a \cdot f(\boldsymbol{v}) = f(a \cdot \boldsymbol{v})$, so $a \cdot \boldsymbol{w} \in \mathrm{im}(f)$. ☺

Injectivity and surjectivity

- A linear map $f : V \to W$ is *surjective*:

$$\forall \boldsymbol{w} \exists \boldsymbol{v}. f(\boldsymbol{v}) = \boldsymbol{w}$$

  if and only if $\text{im}(f) = W$.

- A linear map $f : V \to W$ is *injective*:

$$f(\boldsymbol{v}) = f(\boldsymbol{w}) \implies \boldsymbol{v} = \boldsymbol{w}$$

  if and only if $\ker(f) = \mathbf{0}$.

The kernel as solution space

With this kernel (and image) terminology we can connect some previous concepts.

Theorem

Suppose a linear map $f : V \to W$ has matrix $\mathbf{A}$. Then:

$$
\begin{aligned}
\mathbf{v} \in \ker(f) &\iff f(\mathbf{v}) = \mathbf{0} \\
&\iff \mathbf{A} \cdot \mathbf{v} = \mathbf{0} \\
&\iff \mathbf{v} \textit{ solves a system of homogeneous equations}
\end{aligned}
$$

Moreover, the dimension of the kernel $\dim(\ker(f))$ *is the same as the number of basic solutions of $\mathbf{A}$, that is the number of columns without pivots in the echelon form of $\mathbf{A}$.*

We can learn a lot about a matrix...

- ...by looking at its columns.
- Suppose a linear map f is represented by a matrix $\boldsymbol{A}$ with columns $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$:

$$f(\boldsymbol{w}) = \begin{pmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \\ | & & | \end{pmatrix} \cdot \boldsymbol{w}$$

- Then, $\dim(\mathrm{im}(f))$ is the dimension of the space spanned by $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$
- ...which is the same as the number of pivots in $\boldsymbol{A}$

Kernel-image-dimension theorem (aka. rank-nullity)

Theorem

For a linear map $f : V \to W$ one has:

$$\dim(\ker(f)) + \dim(\mathrm{im}(f)) = \dim(V)$$

Proof: Let $\boldsymbol{A}$ be a matrix that represents f. It has $\dim(V)$ columns. $\dim(\mathrm{im}(f))$ of those are pivots, and the rest correspond to basic solutions to $\boldsymbol{A} \cdot \boldsymbol{x} = \mathbf{0}$, which give a basis for $\ker(f)$.