Matrix Calculations: Diagonalisation, Orthogonality, and Applications

A. Kissinger

Institute for Computing and Information Sciences
Radboud University Nijmegen

Version: autumn 2018

Last time

- Vectors look different in different bases, e.g. for:

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

- we have:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{S}} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}_{\mathcal{C}}$$

Last time

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \qquad \mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

- We can transform bases using basis transformation matrices. Going to standard basis is easy (basis elements are columns):

$$\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad \boldsymbol{T}_{\mathcal{C}\Rightarrow\mathcal{S}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

- ...coming back means taking the inverse:

$$\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} = (\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{C}} = (\boldsymbol{T}_{\mathcal{C}\Rightarrow\mathcal{S}})^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Last time

- The change of basis of a vector is computed by applying the matrix. For example, changing from $\mathcal{S}$ to $\mathcal{B}$ is:

$$\boldsymbol{v}' = \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} \cdot \boldsymbol{v}$$

- The change of basis for a matrix is computed by surrounding it with basis-change matrices.
- Changing from a matrix $\boldsymbol{A}$ in $\mathcal{S}$ to a matrix $\boldsymbol{A}'$ in $\mathcal{B}$ is:

$$\boldsymbol{A}' = \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} \cdot \boldsymbol{A} \cdot \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}}$$

- (Memory aid: look at the first matrix after the equals sign to see what basis transformation you are doing.)

- Many linear maps have their 'own' basis, their *eigenbasis*, which has the property that all basis elements $\boldsymbol{v} \in \mathcal{B}$ do this:

$$\boldsymbol{A} \cdot \boldsymbol{v} = \lambda \boldsymbol{v}$$

- λ is called an eigenvalue, $\boldsymbol{v}$ is called an eigenvector.
- Eigenvalues are computed by solving:

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$$

Outline

Eigenvectors and diagonalisation

Inner products and orthogonality

Wrapping up

Computing eigenvectors

- For an $n \times n$ matrix, the equation $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$ has n solutions, which we'll write as: $\lambda_1, \lambda_2, \ldots, \lambda_n$
- (e.g. a 2×2 matrix involves solving a quadratic equation, which has 2 solutions λ_1 and λ_2)
- For *each* of these solutions, we get a homogeneous system:

$$\underbrace{(\boldsymbol{A} - \lambda_i \boldsymbol{I})}_{\text{matrix}} \cdot \boldsymbol{v}_i = \boldsymbol{0}$$

- Solving this homogeneous system gives us the *associated eigenvector* $\boldsymbol{v}_i$ for the eigenvalue λ_i

Example

- This matrix:

$$\boldsymbol{A} = \begin{pmatrix} 1 & -2 \\ 0 & -1 \end{pmatrix}$$

- Has characteristic polynomial:

$$\det \begin{pmatrix} -\lambda + 1 & -2 \\ 0 & -\lambda - 1 \end{pmatrix} = \lambda^2 - 1$$

- The equation $\lambda^2 - 1 = 0$ has **2 solutions**: $\lambda_1 = 1$ and $\lambda_2 = -1$.

Example

- For $\lambda_1 = 1$, we get a homogeneous system:

$$(\boldsymbol{A} - \lambda_1 \cdot \boldsymbol{I}) \cdot \boldsymbol{v}_1 = \boldsymbol{0}$$

- Computing $(\boldsymbol{A} - (1) \cdot \boldsymbol{I})$:

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

- So, we need to find a *non-zero* solution for:

$$\begin{pmatrix} 0 & -2 \\ 0 & -2 \end{pmatrix} \cdot \boldsymbol{v}_1 = \boldsymbol{0}$$

(just like in lecture 2)

- This works: $\boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Example

- For $\lambda_2 = -1$, we get another homogeneous system:

$$(\boldsymbol{A} - \lambda_2 \cdot \boldsymbol{I}) \cdot \boldsymbol{v}_2 = \mathbf{0}$$

- Computing $(\boldsymbol{A} - (-1) \cdot \boldsymbol{I})$:

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

- So, we need to find a *non-zero* solution for:

$$\begin{pmatrix} 2 & -2 \\ 0 & 0 \end{pmatrix} \cdot \boldsymbol{v}_2 = \mathbf{0}$$

- This works: $\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Example

So, for the matrix $\boldsymbol{A}$, we computed 2 eigenvalue/eigenvector pairs:

$$\lambda_1 = 1, \quad \boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and

$$\lambda_2 = -1, \quad \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Theorem

If the eigenvalues of a matrix $\boldsymbol{A}$ are all different, then their associated eigenvectors form a basis.

Proof. We need to prove the $\boldsymbol{v}_i$ are all linearly independent. Then suppose (for contradiction) that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly *dependent*, i.e.:

$$c_1 \boldsymbol{v}_1 + c_2 \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_n = \boldsymbol{0}$$

for k non-zero coefficients. Then, using that they are eigvectors:

$$\boldsymbol{A} \cdot (c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n) = \boldsymbol{0} \implies \lambda_1 c_1 \boldsymbol{v}_1 + \ldots + \lambda_n c_n \boldsymbol{v}_n = \boldsymbol{0}$$

Suppose $c_j \neq 0$, then subtract $\frac{1}{\lambda_j}$ times 2nd equation from the 1st equation:

$$c_1 \boldsymbol{v}_1 + c_2 \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_n - \frac{1}{\lambda_j}(\lambda_1 c_1 \boldsymbol{v}_1 + \ldots + \lambda_n c_n \boldsymbol{v}_n) = \boldsymbol{0}$$

This has $k - 1$ non-zero coefficients (because all the λ_i's are distinct). Repeat until we have just 1 non-zero coefficient, and we have:

$$c_j \boldsymbol{v}_k = \boldsymbol{0} \implies \boldsymbol{v}_k = \boldsymbol{0}$$

but eigenvectors are always non-zero, so this is a contradiction. □

Changing basis

- Once we have a basis of eigenvectors $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, translating to $\mathcal{B}$ gives us a diagonal matrix, whose diagonal entries are the eigenvalues:

$$\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} \cdot \boldsymbol{A} \cdot \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} = \boldsymbol{D} \qquad \text{where} \qquad \boldsymbol{D} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

- Going the other direction, we can always write $\boldsymbol{A}$ in terms of a diagonal matrix:

$$\boldsymbol{A} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

Definition

For a matrix $\boldsymbol{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ and eigenvectors $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, decomposing $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

is called *diagonalising* the matrix $\boldsymbol{A}$.

Summary: diagonalising a matrix (study this slide!)

We diagonalise a matrix $\boldsymbol{A}$ as follows:

1. Compute each eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_n$ by solving the characteristic polynomial
2. For each eigenvalue, compute the associated eigenvector $\boldsymbol{v}_i$ by solving the homogenious system $(\boldsymbol{A} - \lambda_i \boldsymbol{I}) \cdot \boldsymbol{v}_i = \boldsymbol{0}$.
3. Write down $\boldsymbol{A}$ as the product of three matrices:

$$\boldsymbol{A} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

where:

- $\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}}$ has the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ (in order!) as its columns
- $\boldsymbol{D}$ has the eigenvalues (in the same order!) down its diagonal, and zeroes everywhere else
- $\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$ is the inverse of $\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}}$.

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, ...
- We can represent these computations conveniently using matrix multiplication.

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: $0.2 \cdot 100 + 0.9 \cdot 150 = 20 + 135 = 155$
- If $\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} 100 \\ 150 \end{pmatrix}$, then after one year we have:

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

- After two years we have:

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

Political swingers, part IV

The situation after two years is obtained as:

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

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

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

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

Etc. It looks like $\boldsymbol{P}^{100}$ (or worse, $\lim_{n\to\infty} \boldsymbol{P}^n$) is going to be a real pain to calculate. ...or is it?

Diagonal matrices

- Multiplying lots of matrices together is hard :(
- But multiplying diagonal matrices is easy!

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \cdot \begin{pmatrix} w & 0 & 0 & 0 \\ 0 & x & 0 & 0 \\ 0 & 0 & y & 0 \\ 0 & 0 & 0 & z \end{pmatrix} = \begin{pmatrix} aw & 0 & 0 & 0 \\ 0 & bx & 0 & 0 \\ 0 & 0 & cy & 0 \\ 0 & 0 & 0 & dz \end{pmatrix}$$

- **Strategy:** first *diagonalise* $\boldsymbol{P}$:

$$\boldsymbol{P} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} \qquad \text{where } \boldsymbol{D} \text{ is diagonal}$$

- Then multiply (and see what happens....)

Multiplying diagonalised matrices

- Suppose $\boldsymbol{P} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$, then:

$$\boldsymbol{P} \cdot \boldsymbol{P} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} \cdot \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

- So:

$$\boldsymbol{P} \cdot \boldsymbol{P} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

- and:

$$\boldsymbol{P} \cdot \boldsymbol{P} \cdot \boldsymbol{P} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D} \cdot \boldsymbol{D} \cdot \boldsymbol{D} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

- and so on:

$$\boldsymbol{P}^n = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \boldsymbol{D}^n \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

Political swingers re-revisited, part I

- Suppose we diagonalise the political transition matrix:

$$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}}_{\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}}} \cdot \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 0.7 \end{pmatrix}}_{\boldsymbol{D}} \cdot \underbrace{\frac{1}{3}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}}_{\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}}$$

- Then, raising it to the 10th power is not so hard:

$$\begin{aligned}
\boldsymbol{P}^{10} &= \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0.7 \end{pmatrix}^{10} \cdot \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1^{10} & 0 \\ 0 & 0.7^{10} \end{pmatrix} \cdot \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \\
&\approx \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0.028 \end{pmatrix} \cdot \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \\
&\approx \begin{pmatrix} 0.35 & 0.32 \\ 0.65 & 0.68 \end{pmatrix}
\end{aligned}$$

- We can also compute:

$$
\begin{aligned}
\lim_{n\to\infty} \boldsymbol{P}^n &= \lim_{n\to\infty} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1^n & 0 \\ 0 & 0.7^n \end{pmatrix} \cdot \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \\
&= \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}
\end{aligned}
$$

And more...

- Diagonalisation lets us do lots of things we can normally only do with *numbers* with *matrices* instead
- We already saw raising to a power:

$$\boldsymbol{A}^n = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} \lambda_1^n & 0 & 0 & 0 \\ 0 & \lambda_2^n & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_n^N \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

- We can also do other funky stuff, like take the square root of a matrix:

$$\sqrt{\boldsymbol{A}} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_n} \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

And more...

- Take the square root of a matrix:

$$\sqrt{\boldsymbol{A}} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sqrt{\lambda_n} \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

- (always gives us a matrix where $\sqrt{\boldsymbol{A}} \cdot \sqrt{\boldsymbol{A}} = \boldsymbol{A}$)

And just because they are cool...

- Exponentiate a matrix:

$$e^{\boldsymbol{A}} = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} e^{\lambda_1} & 0 & 0 & 0 \\ 0 & e^{\lambda_2} & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & e^{\lambda_n} \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

(e.g. to solve the Schrödinger equation in quantum mechanics)

- Take the logarithm a matrix:

$$\log(\boldsymbol{A}) = \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} \cdot \begin{pmatrix} \log(\lambda_1) & 0 & 0 & 0 \\ 0 & \log(\lambda_2) & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \log(\lambda_n) \end{pmatrix} \cdot \boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}}$$

(e.g. to compute entropies of quantum states)

Applications: data processing

- **Problem:** suppose we have a HUGE matrix, and we want to know approximately what it looks like
- **Solution:** diagonalise it using its basis $\mathcal{B}$ of eigenvectors...then throw away (= set to zero) all the little eigenvalues:

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & 0 & & 0 \\ \vdots & 0 & \lambda_3 & 0 & \vdots \\ 0 & & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}_{\mathcal{B}} \approx \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & 0 & & 0 \\ \vdots & 0 & 0 & 0 & \vdots \\ 0 & & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{\mathcal{B}}$$

- If there are only a few **big** λ's, and lots of **little** λ's, we get almost the same matrix back
- This is the basic trick used in principle compent analysis (big data) and lossy data compression

Length of a vector

- Each vector $\mathbf{v} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ has a length (aka. norm), written as $\|\mathbf{v}\|$
- This $\|\mathbf{v}\|$ is a non-negative real number: $\|\mathbf{v}\| \in \mathbb{R}, \|\mathbf{v}\| \geq 0$
- Some special cases:
  - $n = 1$: so $\mathbf{v} \in \mathbb{R}$, with $\|\mathbf{v}\| = |\mathbf{v}|$
  - $n = 2$: so $\mathbf{v} = (x_1, x_2) \in \mathbb{R}^2$ and with Pythagoras:

$$\|\mathbf{v}\|^2 = x_1^2 + x_2^2 \qquad \text{and thus} \qquad \|\mathbf{v}\| = \sqrt{x_1^2 + x_2^2}$$

  - $n = 3$: so $\mathbf{v} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and also with Pythagoras:

$$\|\mathbf{v}\|^2 = x_1^2 + x_2^2 + x_3^2 \qquad \text{and thus} \qquad \|\mathbf{v}\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

- In general, for $\mathbf{v} = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|\mathbf{v}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Distance between points

- Assume now we have two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, written as:

$$\boldsymbol{v} = (x_1, \ldots, x_n) \qquad \boldsymbol{w} = (y_1, \ldots, y_n)$$

- What is the distance between the endpoints?
  - commonly written as $d(\boldsymbol{v}, \boldsymbol{w})$
  - again, $d(\boldsymbol{v}, \boldsymbol{w})$ is a non-negative real
- For $n = 2$,

$$d(\boldsymbol{v}, \boldsymbol{w}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} = \|\boldsymbol{v} - \boldsymbol{w}\| = \|\boldsymbol{w} - \boldsymbol{v}\|$$

- This will be used also for other n, so:

$$d(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\|$$

Length is fundamental

- Distance can be obtained from length of vectors
- Angles can also be obtained from length
- Both length of vectors and angles between vectors can be derived from the notion of inner product

Inner product 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 as the real number:

$$\begin{aligned} \langle \boldsymbol{v}, \boldsymbol{w} \rangle &= x_1 y_1 + \cdots + x_n y_n \\ &= \sum_{1 \leq i \leq n} x_i y_i \end{aligned}$$

Note: Length $\|\boldsymbol{v}\|$ can be expressed via inner product:

$$\|\boldsymbol{v}\|^2 = x_1^2 + \cdots + x_n^2 = \langle \boldsymbol{v}, \boldsymbol{v} \rangle, \qquad \text{so} \qquad \|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}.$$

Properties of the inner product

1. The inner product is symmetric in $\boldsymbol{v}$ and $\boldsymbol{w}$:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$$

2. It is linear in $\boldsymbol{v}$:

$$\langle \boldsymbol{v} + \boldsymbol{v}', \boldsymbol{w} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}', \boldsymbol{w} \rangle \qquad \langle a\boldsymbol{v}, \boldsymbol{w} \rangle = a\langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

...and hence also in $\boldsymbol{w}$ (by symmetry):

$$\langle \boldsymbol{v}, \boldsymbol{w} + \boldsymbol{w}' \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w}' \rangle \qquad \langle \boldsymbol{v}, a\boldsymbol{w} \rangle = a\langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

3. And it is positive definite:

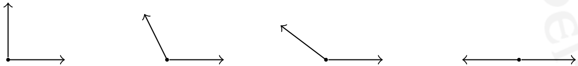
$$\boldsymbol{v} \neq \mathbf{0} \implies \langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0$$

Inner products and angles, part I

For $\boldsymbol{v} = \boldsymbol{w} = (1, 0)$, $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 1$.
As we start to rotate $\boldsymbol{w}$, $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ goes down until 0:

$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 1$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = \frac{4}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = \frac{3}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$

...and then goes to -1:

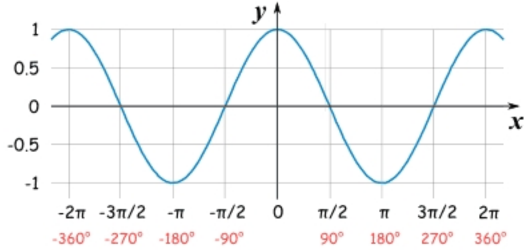


$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{3}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{4}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -1$

...then down to 0 again, then to 1, then repeats...

Cosine

Plotting these numbers vs. the angle between the vectors, we get:



It looks like $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ depends on the cosine of the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$.

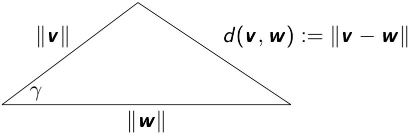
- In fact, if $\|v\| = \|w\| = 1$, it *is* true that $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \cos \gamma$.
- For the general equation, we need to divide by their lengths:

$$\cos(\gamma) = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\| \, \|\boldsymbol{w}\|}$$

- Remember this equation!

Inner products and angles, part II

Proof (sketch). For 2 any two vectors, we can make a triangle like this:



Then, we apply the *cosine rule* from trig to get:

$$\cos(\gamma) \;=\; \frac{a^2 + b^2 - c^2}{2ab} \;=\; \frac{\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 - \|\boldsymbol{v} - \boldsymbol{w}\|^2}{2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|}$$

...then after expanding the definition of $\|.\|$ and some work we get:

$$\cos(\gamma) = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|}$$

Examples

- What is the angle between $(1,1)$ and $(-1,-1)$?

$$\cos\gamma = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} = \frac{-2}{\sqrt{2}\cdot\sqrt{2}} = \frac{-2}{2} = -1 \qquad \Longrightarrow \qquad \gamma = \pi$$

- What is the angle between $(1,0)$ and $(1,1)$?

$$\cos\gamma = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} = \frac{1}{1\cdot\sqrt{2}} = \frac{1}{\sqrt{2}} \qquad \Longrightarrow \qquad \gamma = \frac{\pi}{4}$$

- What is the angle between $(1,0)$ and $(0,1)$?

$$\cos\gamma = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} = \frac{0}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} = 0 \qquad \Longrightarrow \qquad \gamma = \frac{\pi}{2}$$

Orthogonality

Definition

Two vectors $\mathbf{v}, \mathbf{w}$ are called orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. This is written as $\mathbf{v} \perp \mathbf{w}$.

Explanation: orthogonality means that the cosine of the angle between the two vectors is 0; hence they are perpendicular.

Example

Which vectors $(x, y) \in \mathbb{R}^2$ are orthogonal to $(1, 1)$?

Examples, are $(1, -1)$ or $(-1, 1)$, or more generally $(x, -x)$.

This follows from an easy computation:

$$\langle (x, y), (1, 1) \rangle = 0 \iff x + y = 0 \iff y = -x.$$

Orthogonality and independence

Lemma

Call a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of **non-zero** *vectors orthogonal if every pair of different vectors is orthogonal.*

1. *orthogonal vectors are always independent,*
2. *independent vectors are not always orthogonal.*

Proof: The second point is easy: $(1, 1)$ and $(1, 0)$ are independent, but not orthogonal

Orthogonality and independence (cntd)

(Orthogonality $\Longrightarrow$ Independence): assume $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is orthogonal and $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = \mathbf{0}$. Then for each $i \leq n$:

$$\begin{aligned} 0 &= \langle \mathbf{0}, \boldsymbol{v}_i \rangle \\ &= \langle a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n, \boldsymbol{v}_i \rangle \\ &= \langle a_1\boldsymbol{v}_1, \boldsymbol{v}_i \rangle + \cdots + \langle a_n\boldsymbol{v}_n, \boldsymbol{v}_i \rangle \\ &= a_1\langle \boldsymbol{v}_1, \boldsymbol{v}_i \rangle + \cdots + a_n\langle \boldsymbol{v}_n, \boldsymbol{v}_i \rangle \\ &= a_i\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle \qquad \text{since } \langle \boldsymbol{v}_j, \boldsymbol{v}_i \rangle = 0 \text{ for } j \neq i \end{aligned}$$

But since $\boldsymbol{v}_i \neq 0$ we have $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle \neq 0$, and thus $a_i = 0$.
This holds for each i, so $a_1 = \cdots = a_n = 0$, and we have proven independence. ☺

Orthogonal and orthonormal bases

Definition

A basis $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of a vector space with an inner product is called:

1. orthogonal if $\mathcal{B}$ is an orthogonal set: $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ if $i \neq j$
2. orthonormal if it is orthogonal and $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = \|\boldsymbol{v}_i\| = 1$, for each i

Example

The standard basis $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \cdots, (0, \cdots, 0, 1)$ is an orthonormal basis of $\mathbb{R}^n$.

From independence to orthogonality

- Not every basis is an orthonormal basis:

$$\boxed{\text{Orthonormal basis}} \; \underset{\not\Longleftarrow}{\Longrightarrow} \; \boxed{\text{Basis}}$$

- But, by taking linear linear combinations of basis vectors, we can transform a basis into a (better) orthonormal basis:

$$\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \quad \mapsto \quad \mathcal{B}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$$

- Making basis vectors *normalised* is easy:

$$\mathbf{v}_i \quad \mapsto \quad \mathbf{w}_i := \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$$

- Making vectors orthogonal is also always possible, using a procedure called *Gram-Schmidt orthogonalisation*.

In summary

- The inner product gives us a means to compute the lengths of vectors:

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

- It also lets us compute the angles between vectors:

$$\cos(\gamma) = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\| \, \|\boldsymbol{w}\|}$$

- $\Rightarrow$ vectors with very large inner product are very close to pointing the same direction (because $\cos(0) = 1$)
- $\Rightarrow$ vectors with very small inner product are very close to orthogonal (because $\cos(\pi/2) = 0$)
- $\Rightarrow$ inner products measure *how similar* two vectors are.

Application: Computational linguistics

Computational linguistics = teaching computers to read

- **Example:** I have two words, and I want a program that tells me how "similar" the two words are, e.g.

$$\begin{aligned} \text{nice} + \text{kind} &\Rightarrow 95\% \text{ similar} \\ \text{dog} + \text{cat} &\Rightarrow 61\% \text{ similar} \\ \text{dog} + \text{xylophone} &\Rightarrow 0.1\% \text{ similar} \end{aligned}$$

- **Applications:** thesaurus, smart web search, translation, ...
- Dumb solution: ask a whole bunch of people to rate similarity and make a big database
- Smart solution: use *distributional semantics*

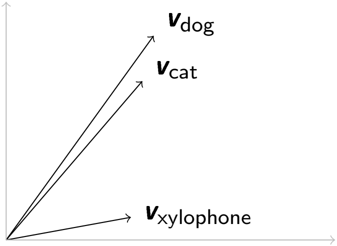
Meaning vectors

"You shall know a word by the company it keeps."
– J. R. Firth

- Pick about 500-1000 words ($\boldsymbol{v}_{\text{cat}}$, $\boldsymbol{v}_{\text{boy}}$, $\boldsymbol{v}_{\text{sandwich}}$ ...) to act as "basis vectors"
- Build up a meaning vector for each word, e.g. "dog", by scanng a whole lot of text
- Every time "dog" occurs within, say 200 words of a basis vector, add that basis vector. Soon we'll have:

$$\boldsymbol{v}_{\text{dog}} = 2308198 \cdot \boldsymbol{v}_{\text{cat}} + 4291 \cdot \boldsymbol{v}_{\text{boy}} + 4 \cdot \boldsymbol{v}_{\text{sandwich}} + \cdots$$

- Similar words cluster together:



- ...while dissimilar words drift apart.We can measure this by:

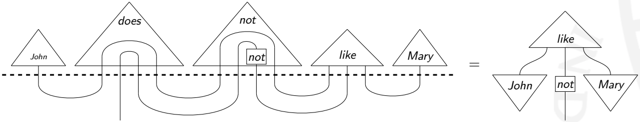
$$\frac{\langle \mathbf{v}_{\text{dog}}, \mathbf{v}_{\text{cat}} \rangle}{\|\mathbf{v}_{\text{dog}}\| \, \|\mathbf{v}_{\text{cat}}\|} = 0.953 \qquad \frac{\langle \mathbf{v}_{\text{dog}}, \mathbf{v}_{\text{xylophone}} \rangle}{\|\mathbf{v}_{\text{dog}}\| \, \|\mathbf{v}_{\text{xylophone}}\|} = 0.001$$

- Search engines do something very similar. Learn more in the course on Information Retrieval.

Distributional Semantics

- This works very well, but also has weaknesses (e.g. meanings of whole sentences, ambiguous words)
- This can be improved by incorporating other kinds of semantics:

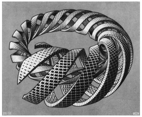
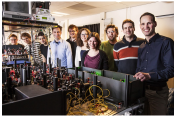
distributional + **co**mpositional + **cat**egorical



= **DisCoCat**

About linear algebra

- Linear algebra forms a coherent body of mathematics . . .
- involving elementary algebraic and geometric notions
    - systems of equations and their solutions
    - vector spaces with bases and linear maps
    - matrices and their operations (product, inverse, determinant)
    - inner products and distance
- . . . together with various calculational techniques
    - the most important/basic ones you learned in this course
    - they are used all over the place: mathematics, physics, engineering, linguistics...

About the exam, part I

- Closed book
  - Simple '4-function' calculators are allowed (but not necessary)
  - phones, graphing calculators, etc. are NOT allowed
- Questions are in line with exercises from assignments
- In principle, slides contain all necessary material
  - LNBS lecture notes have extra material for practice
  - wikipedia also explains a lot
- Theorems, definitions, etc:
  - are needed to understand the theory
  - are needed to answer the questions
  - their proofs are not required for the exam (but do help understanding)
    - need *not* be reproducable literally
    - but help you to understand questions

About the exam, part II

Calculation rules (or formulas) must be known by heart for:

1. solving (non)homogeneous equations, echelon form
2. linearity, independence, matrix-vector multiplication
3. matrix multiplication & inverse, change-of-basis matrices
4. eigenvalues, eigenvectors and determinants
5. inner products, distance, length, angle, orthogonality

About the exam, part III

- Questions are formulated in English
  - you may choose to answer in Dutch or English
- Give intermediate calculation results
  - just giving the outcome (say: 68) yields no points when the answer should be 67
- Write legibly, and explain what you are doing
  - giving explanations forces yourself to think systematically
  - mitigates calculation mistakes
- Perform checks yourself, whenever possible, e.g.
  - solutions of equations
  - inverses of matrices,
  - orthogonality of vectors, etc.

Finally . . .

Practice, practice, practice!

(so that you can rely on skills, not on luck)

Some practical issues (Autumn 2018)

- Exam: Tuesday, October 30, 8:30–10:30 in HAL 2. (Extra time: 8:30-11:00, HG00.108)
- Vragenuur: there will be a Q&A session next week. Friday, 26 October. 13:30-15:15 in MERC1 00.28
- How we compute the final grade g for the course
    - Your exam grade e, which should be ≥ 5,
    - Your average assignment grade a
    - Final grade is: $e + \frac{a}{10}$, rounded to the nearest half (except 5.5).

Final request

- Fill out the enquete form for *Matrixrekenen*, IPC017, when invited to do so.
- Any constructive feedback is highly appreciated.

And good luck with the preparation & exam itself!
Start now!