Matrix Calculations: Inner Products & Orthogonality

A. Kissinger

Institute for Computing and Information Sciences
Radboud University Nijmegen

Version: autumn 2017

Outline

Length of a vector

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

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

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

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

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

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

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:

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

Inner products via matrix transpose

Matrix transposition

For an $m \times n$ matrix $\boldsymbol{A}$, the transpose $\boldsymbol{A}^T$ is the $n \times m$ matrix $\boldsymbol{A}$ obtained by mirroring in the diagonal:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \vdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ & \vdots & \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}$$

In other words, the rows of $\boldsymbol{A}$ become the columns of $\boldsymbol{A}^T$.

The inner product of $\boldsymbol{v} = (x_1, \ldots, x_n), \boldsymbol{w} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ is then a matrix product:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = x_1 y_1 + \cdots + x_n y_n = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{v}^T \cdot \boldsymbol{w}.$$

Properties of the inner product

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

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

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

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

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

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

3. And it is positive definite:

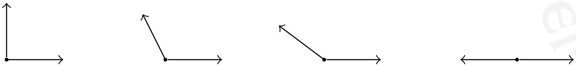
$$\mathbf{v} \neq \mathbf{0} \implies \langle \mathbf{v}, \mathbf{v} \rangle > 0$$

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:

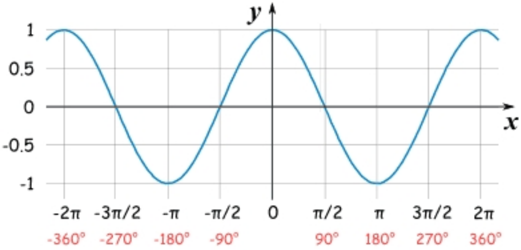


$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{3}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{4}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -1$

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

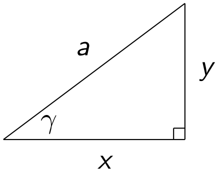
Cosine

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



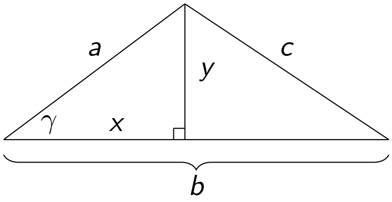
It looks like $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ depends on the cosine of the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$. Let's prove it!

Recall: definition of cosine



$$\cos(\gamma) = \frac{x}{a} \qquad \Longrightarrow \qquad x \;=\; a\cos(\gamma)$$

The cosine rule



Claim: $\cos(\gamma) = \dfrac{a^2 + b^2 - c^2}{2ab}$

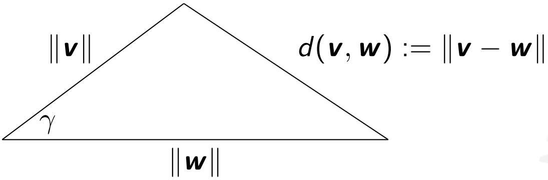
Proof: We have three equations to play with:

$$x^2 + y^2 = a^2 \qquad (b - x)^2 + y^2 = c^2 \qquad x = a\cos(\gamma)$$

...do the math. ☺

Inner products and angles, part II

Translating this to something about vectors:



gives:

$$\cos(\gamma) = \frac{\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 - \|\boldsymbol{v} - \boldsymbol{w}\|^2}{2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|}$$

Let's clean this up...

Inner products and angles, part II

Starting from the cosine rule:

$$\begin{aligned}
\cos(\gamma) &= \frac{\|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 - \|\boldsymbol{v} - \boldsymbol{w}\|^2}{2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} \\
&= \frac{x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 - (x_1 - y_1)^2 - \cdots - (x_n - y_n)^2}{2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} \\
&= \frac{2x_1y_1 + \cdots + 2x_ny_n}{2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} \\
&= \frac{x_1y_1 + \cdots + x_ny_n}{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|} \\
&= \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|}
\end{aligned}$$

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

Thus, angles between vectors are expressible via the inner product (since $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$).

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 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is orthogonal and $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = 0$. Then for each $i \leq n$:

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

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 $\mathbb{R}^n$.

Orthonormal basis transformations

- Orthonormal bases are very handy! Example: basis transformations.
- For any basis $\mathcal{B}$, the matrix $\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}}$ is easy to compute: it has the vectors in $\mathcal{B}$ as its columns.
- Normally, $\boldsymbol{T}_{\mathcal{S}\Rightarrow\mathcal{B}} := (\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{-1}$ is a pain to compute, but $(\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{T}$ is also easy: it has the vectors in $\mathcal{B}$ as its rows
- Now, if $\mathcal{B}$ is an orthonormal basis, a miracle occurs:

$$(\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{T} \cdot \boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}} = \begin{pmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle & \langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle & \cdots & \langle \boldsymbol{v}_1, \boldsymbol{v}_n\rangle \\ \langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle & \cdots & \langle \boldsymbol{v}_2, \boldsymbol{v}_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{v}_n, \boldsymbol{v}_1\rangle & \langle \boldsymbol{v}_n, \boldsymbol{v}_2\rangle & \cdots & \langle \boldsymbol{v}_n, \boldsymbol{v}_n\rangle \end{pmatrix} = \boldsymbol{I}$$

- So, $(\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{-1} = (\boldsymbol{T}_{\mathcal{B}\Rightarrow\mathcal{S}})^{T}$!

From independence to orthogonality

- Not every basis is an orthonormal basis:

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

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

$$\mathcal{B} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \quad \mapsto \quad \mathcal{B}' = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$$

- Making basis vectors *normalised* is easy:

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

- But first they should be orthogonal, which we can accomplish using Gram-Schmidt orthogonalisation

Making vectors orthogonal

- Suppose we have two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ which are independent, but not orthogonal
- We want to make a new orthogonal pair of vectors $\boldsymbol{w}_1, \boldsymbol{w}_2$ which span the same space.
- We do it one vector at a time, starting with $\boldsymbol{w}_1 = \boldsymbol{v}_1$
- Then $\boldsymbol{v}_2$ has a "bit of $\boldsymbol{w}_1$" in it:

$$\boldsymbol{v}_2 = \lambda \boldsymbol{w}_1 + \underbrace{\cdots\cdots\cdots}_{\text{stuff that is orthogonal to } \boldsymbol{w}_1}$$

- So lets take it out! Let $\boldsymbol{w}_2 := \boldsymbol{v}_2 - \lambda \boldsymbol{w}_1$

Making vectors orthogonal

$$\boldsymbol{w}_2 := \boldsymbol{v}_2 - \lambda \boldsymbol{w}_1$$

- The only thing we need to do is find λ. Here's what we want:

$$0 = \langle \boldsymbol{w}_2, \boldsymbol{w}_1 \rangle = \langle \boldsymbol{v}_2 - \lambda \boldsymbol{w}_1, \boldsymbol{e}_1 \rangle = \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle - \lambda \langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle$$

$$\Longrightarrow \quad \lambda = \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \quad \Longrightarrow \quad \boldsymbol{w}_2 = \boldsymbol{v}_2 - \underbrace{\frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1}_{\text{the `projection of } \boldsymbol{v}_2 \text{ onto } \boldsymbol{w}_1\text{'}}$$

Making vectors orthogonal

- The process continues...
- We make $\boldsymbol{w}_3$ by starting with $\boldsymbol{v}_3$, and removing the stuff that is not orthogonal to $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$:

$$\boldsymbol{w}_3 = \boldsymbol{v}_3 - \underbrace{\frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1}_{\text{proj. of } \boldsymbol{v}_3 \text{ onto } \boldsymbol{w}_1} - \underbrace{\frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_2 \rangle}{\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle} \boldsymbol{w}_2}_{\text{proj. of } \boldsymbol{v}_3 \text{ onto } \boldsymbol{w}_2}$$

- and so on...

$$\boldsymbol{w}_4 = \boldsymbol{v}_4 - \underbrace{\frac{\langle \boldsymbol{v}_4, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1}_{\text{proj. of } \boldsymbol{v}_4 \text{ onto } \boldsymbol{w}_1} - \underbrace{\frac{\langle \boldsymbol{v}_4, \boldsymbol{w}_2 \rangle}{\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle} \boldsymbol{w}_2}_{\text{proj. of } \boldsymbol{v}_4 \text{ onto } \boldsymbol{w}_2} - \underbrace{\frac{\langle \boldsymbol{v}_4, \boldsymbol{w}_3 \rangle}{\langle \boldsymbol{w}_3, \boldsymbol{w}_3 \rangle} \boldsymbol{w}_3}_{\text{proj. of } \boldsymbol{v}_4 \text{ onto } \boldsymbol{w}_3}$$

Gram-Schmidt orthogonalisation

Start with an independent set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of vectors.

Make them orthogonal one at a time:

$$\begin{aligned}
\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} &\Rightarrow \{\boldsymbol{w}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \\
&\Rightarrow \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{v}_n\} \\
&\qquad \cdots \\
&\Rightarrow \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n\}
\end{aligned}$$

...where each $\boldsymbol{w}_i$ depends only on $\boldsymbol{v}_i$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{i-1}$, i.e. the orthogonal vectors we have made already.

Gram-Schmidt orthogonalisation, part I

1. Starting point: independent set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of vectors
2. Take $\boldsymbol{w}_1 = \boldsymbol{v}_1$
3. Take $\boldsymbol{w}_2 = \boldsymbol{v}_2 - \dfrac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1$

   This gives an orthogonal vector. We can check:

$$\begin{aligned}
\langle \boldsymbol{w}_2, \boldsymbol{w}_1 \rangle &= \left\langle \boldsymbol{v}_2 - \tfrac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1, \boldsymbol{w}_1 \right\rangle \\
&= \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle - \left\langle \tfrac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1, \boldsymbol{w}_1 \right\rangle \\
&= \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle - \tfrac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle \\
&= \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle - \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle \\
&= 0
\end{aligned}$$

Gram-Schmidt orthogonalisation, part II

4. In general, $\boldsymbol{w}_i = \boldsymbol{v}_i - \frac{\langle \boldsymbol{v}_i, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle}\boldsymbol{w}_1 - \cdots - \frac{\langle \boldsymbol{v}_i, \boldsymbol{w}_{i-1} \rangle}{\langle \boldsymbol{w}_{i-1}, \boldsymbol{w}_{i-1} \rangle}\boldsymbol{w}_{i-1}$

   By essentially the same reasoning as before one shows:

   $$\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle = 0, \qquad \text{for all } j < i.$$

5. Result: orthogonal set of vectors $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$.

Gram-Schmidt orthogonalisation: example I

- Take $\boldsymbol{v}_1 = (1, -1)$ and $\boldsymbol{v}_2 = (2, 1)$ in $\mathbb{R}^2$.
- Clearly not orthogonal! $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 1$
- Lets fix that. Let $\boldsymbol{w}_1 := \boldsymbol{v}_1$ and:

$$\begin{aligned} \boldsymbol{w}_2 &= \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \end{pmatrix} \end{aligned}$$

- Bam! $\langle \boldsymbol{w}_1, \boldsymbol{w}_2 \rangle = 0$

Gram-Schmidt orthogonalisation: example II

- Take in $\mathbb{R}^4$, $\boldsymbol{v}_1 = (0,1,2,1)$, $\boldsymbol{v}_2 = (0,1,3,1)$, $\boldsymbol{v}_3 = (1,1,1,0)$
- $\boldsymbol{w}_1 = \boldsymbol{v}_1 = (0,1,2,1)$; then $\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle = 1\cdot 1 + 2 \cdot 2 + 1 \cdot 1 = 6$.
- $$\begin{aligned} \boldsymbol{w}_2 &= \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 \\ &= (0,1,3,1) - \tfrac{1\cdot 1 + 3\cdot 2 + 1 \cdot 1}{6}(0,1,2,1) \\ &= (0,1,3,1) - \tfrac{8}{6}(0,1,2,1) = (0, -\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{1}{3}) \quad \checkmark \end{aligned}$$

  We prefer to take: $\boldsymbol{w}_2 = (0,-1,1,-1)$; then $\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle = 3$.
- $$\begin{aligned} \boldsymbol{w}_3 &= \boldsymbol{v}_3 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_2 \rangle}{\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle} \boldsymbol{w}_2 \\ &= \cdots = (1, \tfrac{1}{2}, 0, -\tfrac{1}{2}) \quad \checkmark \end{aligned}$$

  We can change it into $\boldsymbol{w}_3 = (2,1,0,-1)$, for convenience.

Making an orthonormal basis

Definition

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

1. orthogonal if 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 $\|\boldsymbol{v}_i\| = 1$, for each i

By Gram-Schmidt each basis can be made orthogonal (first), and then orthonormal by replacing $\boldsymbol{v}_i$ by $\frac{1}{\|\boldsymbol{v}_i\|}\boldsymbol{v}_i$.

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*

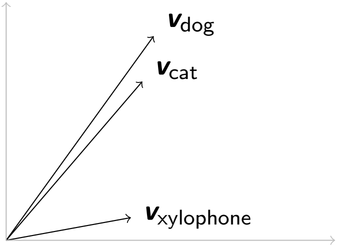
Meaning vectors

"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:

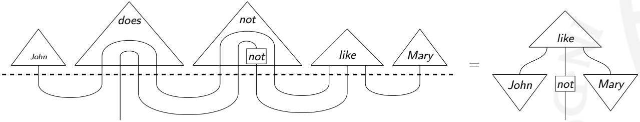
$$\frac{\langle \boldsymbol{v}_{\text{dog}}, \boldsymbol{v}_{\text{cat}} \rangle}{\|\boldsymbol{v}_{\text{dog}}\| \, \|\boldsymbol{v}_{\text{cat}}\|} = 0.953 \qquad \frac{\langle \boldsymbol{v}_{\text{dog}}, \boldsymbol{v}_{\text{xylophone}} \rangle}{\|\boldsymbol{v}_{\text{dog}}\| \, \|\boldsymbol{v}_{\text{xylophone}}\|}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

- 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:

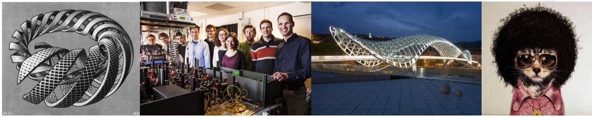
distributional + **co**mpositional + **cat**egorical



= **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, propositions, lemmas:
  - 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, Gram-Schmidt orthogonalisation

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 2017)

- Exam: Monday, November 6, 8:30–11:30 in HAL 1.
  (Extra time: 8:30-12:00, HG00.062)
- Vragenuur: there will be a Q&A session next week. Thursday, 2 November. 8:45-10:30 in HG00.062
- How we compute the final grade g for the course
  - Your exam grade e
  - 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!