Matrix Calculations: Inner Products & Orthogonality

A. Kissinger (and H. Geuvers)

Institute for Computing and Information Sciences
Radboud University Nijmegen

Version: spring 2016

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 $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, written as:

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

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

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

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

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

Length is fundamental

- Distance can be obtained from length of vectors
- Interestingly, also angles can be obtained from length!
- Both length of vectors and angles between vectors can de 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:

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

Inner products via matrix transpose

Recall matrix transposition

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

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

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 $\boldsymbol{v}$ 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:

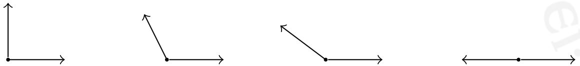
$$\boldsymbol{v} \neq \boldsymbol{0} \implies \langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0$$

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:

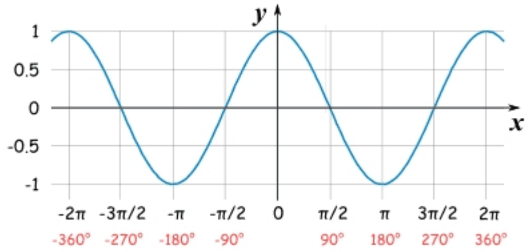


$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{3}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\frac{4}{5}$ $\quad \langle \boldsymbol{v}, \boldsymbol{w} \rangle = -1$

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

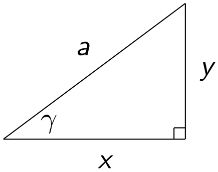
Cosine

After a bit of staring, we see that $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ depends on the cosine of the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$:



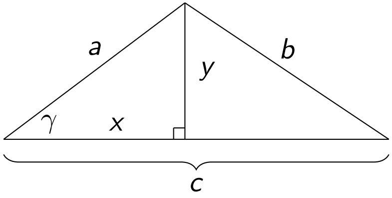
Let's prove it!

Recall: definition of cosine



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

The cosine rule



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

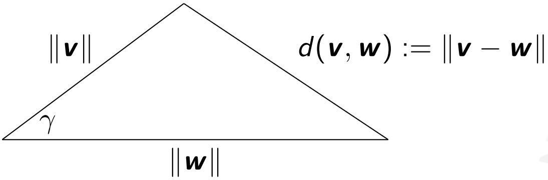
Proof: We have three equations to play with:

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

...lets 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:

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

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

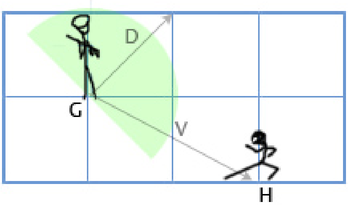
Thus, angles between vectors are expressible via the inner product (since $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$).

Linear algebra in gaming, part I

- Linear algebra plays an important role in game visualisation
- Here: simple illustration, borrowed from blog.wolfire.com
  (More precisely: http://blog.wolfire.com/2009/07/linear-algebra-for-game-developers-part-2)
- Recall: cosine cos function is positive on angles between -90 and +90 degrees.

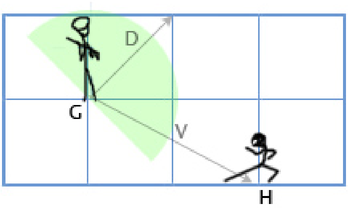
Linear algebra in gaming, part II

- Consider a guard $\boldsymbol{G}$ and hiding ninja $\boldsymbol{H}$ in:



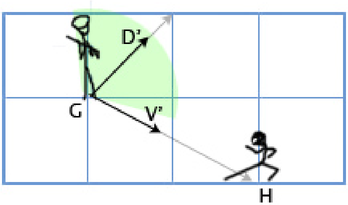
- The **guard** is at position $(1,1)$, facing in direction $\boldsymbol{D} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, with a 180 degree field of view
- The **ninja** is at $(3,0)$. Is he in sight?

Linear algebra in gaming, part III



- The vector from $\boldsymbol{G}$ to $\boldsymbol{H}$ is: $\boldsymbol{V} = \begin{pmatrix} 3 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$
- The angle γ between $\boldsymbol{D}$ and $\boldsymbol{V}$ must be between -90 and +90
- Hence we must have: $\cos(\gamma) = \frac{\langle \boldsymbol{D}, \boldsymbol{V} \rangle}{\|\boldsymbol{D}\| \cdot \|\boldsymbol{V}\|} \geq 0$
- Since $\|\boldsymbol{D}\| \geq 0$ and $\|\boldsymbol{V}\| \geq 0$, it suffices to have: $\langle \boldsymbol{D}, \boldsymbol{V} \rangle \geq 0$
- Well, $\langle \boldsymbol{D}, \boldsymbol{V} \rangle = 1 \cdot 2 + 1 \cdot -1 = 1$. Hence $\boldsymbol{H}$ is within sight!

Linear algebra in gaming, part IV



- Now what if the guard's field of view is 60 degrees?
- Inbetween -30 and +30 degrees we have $\cos(\gamma) \geq \frac{1}{2}\sqrt{3} \sim 0.87$
- The cosine of the actual angle γ between $\boldsymbol{D}$ and $\boldsymbol{V}$ is:

$$\cos(\gamma) = \frac{\langle \boldsymbol{D}, \boldsymbol{V} \rangle}{\|\boldsymbol{D}\| \cdot \|\boldsymbol{V}\|} = \frac{1 \cdot 2 + 1 \cdot -1}{\sqrt{1^2 + 1^2} \cdot \sqrt{2^2 + (-1)^2}}$$
$$= \frac{1}{\sqrt{2} \cdot \sqrt{5}} \sim 0.31 \; < 0.87$$

- ***H* is now out of view!** (the angle $\gamma = \cos^{-1}(0.31) = 72$ degr.)

Orthogonality

Definition

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

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:

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

Pythagoras law, via inner products

Theorem

For orthogonal vectors $\boldsymbol{v}, \boldsymbol{w}$,

$$\|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2$$

Proof: If $\boldsymbol{v} \perp \boldsymbol{w}$, that is, $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$, then:

$$\begin{aligned}
\|\boldsymbol{v} - \boldsymbol{w}\|^2 &= \langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{v} - \boldsymbol{w} \rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{v} - \boldsymbol{w} \rangle + \langle -\boldsymbol{w}, \boldsymbol{v} - \boldsymbol{w} \rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{v} \rangle - \langle \boldsymbol{v}, \boldsymbol{w} \rangle - \langle \boldsymbol{w}, \boldsymbol{v} \rangle + \langle \boldsymbol{w}, \boldsymbol{w} \rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{v} \rangle - 0 - 0 + \langle \boldsymbol{w}, \boldsymbol{w} \rangle \\
&= \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2
\end{aligned}$$

☺

Orthogonality and independence

Lemma

Call a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *of* **non-zero** *vectors orthogonal if they are pairwise orthogonal.*

1. *such an orthogonal collection consists of independent vectors*
2. *independent vectors need not be 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.

From independence to orthogonality

- We have seen that, for a set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of non-zero vectors:

$$\boxed{\text{Orthogonality}} \; \overset{\Longrightarrow}{\underset{\not\Longleftarrow}{}} \; \boxed{\text{Independence}}$$

- But we can transform an independent set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of vectors into an orthogonal set $\{\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_n\}$.
- This procedure is called Gram-Schmidt orthogonalisation

Making vectors orthogonal

- Suppose we have two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ which are independent, but not orthogonal
- Then $\boldsymbol{v}_2$ has a "bit of $\boldsymbol{v}_1$" in it:

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

- So lets take it out! Let $\boldsymbol{v}_2' := \boldsymbol{v}_2 - \lambda \boldsymbol{v}_1$
- The only thing we need to do is find λ. Here's what we want:

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

$$\implies \quad \lambda = \frac{\langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \quad \implies \quad \boldsymbol{v}_2' = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1$$

Gram-Schmidt orthogonalisation: the idea

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

Make them orthogonal one at a time:

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

...where each $\boldsymbol{v}_i'$ depends only on $\boldsymbol{v}_i$ and $\boldsymbol{v}_1', \ldots, \boldsymbol{v}_{i-1}'$, 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 $\boldsymbol{v}_1' = \boldsymbol{v}_1$
3. Take $\boldsymbol{v}_2' = \boldsymbol{v}_2 - \dfrac{\langle \boldsymbol{v}_2, \boldsymbol{v}_1' \rangle}{\langle \boldsymbol{v}_1', \boldsymbol{v}_1' \rangle} \boldsymbol{v}_1'$

   This gives an orthogonal vector:

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

Gram-Schmidt orthogonalisation, part II

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

   By essentially the same reasoning as before one shows:

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

5. Result: orthogonal set $\{\boldsymbol{v}_1', \ldots, \boldsymbol{v}_n'\}$.

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 $\boldsymbol{v}_1' := \boldsymbol{v}_1$ and:

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

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

Gram-Schmidt orthogonalisation: example II

- Take in $\mathbb{R}^4$, $\boldsymbol{v}_1 = (0,1,2,1)$, $\boldsymbol{v}_2 = (0,1,3,1)$, $\boldsymbol{v}_3 = (1,1,1,0)$
- $\boldsymbol{v}_1' = \boldsymbol{v}_1 = (0,1,2,1)$; then $\langle \boldsymbol{v}_1', \boldsymbol{v}_1' \rangle = 1\cdot 1 + 2\cdot 2 + 1\cdot 1 = 6$.
- $$\begin{aligned} \boldsymbol{v}_2' &= \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{v}_1' \rangle}{\langle \boldsymbol{v}_1', \boldsymbol{v}_1' \rangle} \boldsymbol{v}_1' \\ &= (0,1,3,1) - \tfrac{1\cdot 1+3\cdot 2+1\cdot 1}{6}(0,1,2,1) \\ &= (0,1,3,1) - \tfrac{8}{6}(0,1,2,1) \ = \ (0,-\tfrac{1}{3},\tfrac{1}{3},-\tfrac{1}{3}) \quad \checkmark \end{aligned}$$
  We prefer to take: $\boldsymbol{v}_2' = (0,-1,1,-1)$; then $\langle \boldsymbol{v}_2', \boldsymbol{v}_2' \rangle = 3$.
- $$\begin{aligned} \boldsymbol{v}_3' &= \boldsymbol{v}_3 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{v}_1' \rangle}{\langle \boldsymbol{v}_1', \boldsymbol{v}_1' \rangle} \boldsymbol{v}_1' - \frac{\langle \boldsymbol{v}_3, \boldsymbol{v}_2' \rangle}{\langle \boldsymbol{v}_2', \boldsymbol{v}_2' \rangle} \boldsymbol{v}_2' \\ &= \cdots \ = \ (1,\tfrac{1}{2},0,-\tfrac{1}{2}) \quad \checkmark \end{aligned}$$
  We can change it into $\boldsymbol{v}_3' = (2,1,0,-1)$, for convenience.

Orthogonal and orthonormal bases

Definition

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

1. orthogonal if B is an orthogonal set: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ if $i \neq j$
2. orthonormal if it is orthogonal and $\|\mathbf{v}_i\| = 1$, 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 $\mathbb{R}^n$.

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

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*

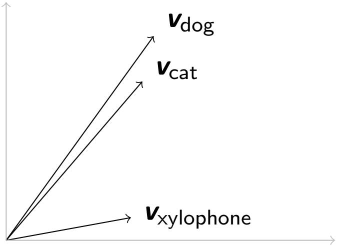
Meaning vectors

"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:

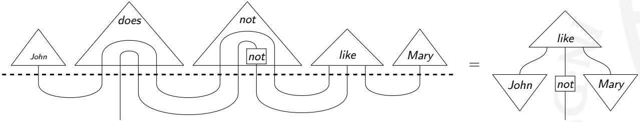
$$\frac{\langle \mathbf{v}_{\text{dog}}, \mathbf{v}_{\text{cat}} \rangle}{\|\mathbf{v}_{\text{dog}}\| \, \|\mathbf{v}_{\text{cat}}\|} = 0.953 \qquad \frac{\langle \mathbf{v}_{\text{dog}}, \mathbf{v}_{\text{xylophone}} \rangle}{\|\mathbf{v}_{\text{dog}}\| \, \|\mathbf{v}_{\text{xylophone}}\|} = 0.001$$

- 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:

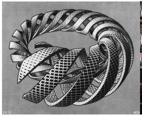
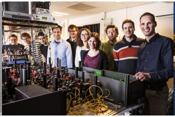
distributional + **co**mpositional + **cat**egorical



= **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; no calculator / smart phone / $\cdots$ allowed
- Questions are in line with exercises from assignments
- In principle, slides contain all necessary material
  - LNBS is background material
  - 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 normal form
2. linearity, independence, matrix-vector multiplication, kernel & image
3. matrix multiplication & inverse, change-of-basis matrices
4. eigenvalues, eigenvectors and determinants
5. inner products, distance, length, angle, orthogonality, Gram-Schmidt orthogonalisation

5 will be on the exam. Sample problems/solutions available on the website (see exercise page). Do these!

No final werkcollege. For questions: email your TA or ask at the vragenuur.

About the exam, part III

- Questions are formulated in English
  - you may choose to answer in Dutch or English (no other languages!)
- 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 (Spring 2016)

- Exam: Monday, April 4, 12:30–15:30 in HAL 2.
- Vragenuur: Friday April 1, 13:45-15.30 in HG00.304, about course material & old exam (will appear on webpage).
- How we compute the final grade g for the course
  - Your exam grade e, which should be ≥ 5,
  - Your average assignment grade a, which should be ≥ 5.
  - We compute $e + \frac{a}{10}$ and round it to the nearest half (except 5.5).

Some more practical issues (Spring 2016)

Students who do the exam for the third (or more) time:

- You should register 1 week before the exam.
- Bring your filled-in registration form (after this lecture or to my office: Mercator 1, 03.02) and I will sign it.
- Next, go to the student desk of FNWI and deliver your form

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!