Matrix Calculations: Orthogonality and Projections

H. Geuvers

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

Version: spring 2015
Outline

Orthogonal projections

Application: search engines

Wrapping up
Orthogonal projections
Application: search engines
Wrapping up

Radboud University Nijmegen

Remember the inner product for vector spaces

**Definition**

A vector space $V$ has an **inner product** if there is a function:

$$ V \times V \xrightarrow{\langle -,- \rangle} \mathbb{R} $$

satisfying:

1. $\langle v, v \rangle \geq 0$
2. $\langle v, v \rangle = 0$ if and only if $v = 0$
3. $\langle v, w \rangle = \langle w, v \rangle$
4. $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$ (similarly in $w$, by 3)
5. $\langle av, w \rangle = a \langle v, w \rangle$ (and similarly in $w$, by 3)

Given such inner product, we define length, distance and angle:

$$ \| v \| = \sqrt{\langle v, v \rangle} \quad d(v, w) = \| v - w \| \quad \cos(\gamma) = \frac{\langle v, w \rangle}{\| v \| \| w \|}. $$

The **standard inner product** $\langle -,- \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is indeed an inner product.
Orthogonality

Definition

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

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:

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

Lemma

1. $v \perp w$ implies $w \perp v$
2. $v \perp 0$, i.e. $\langle v, 0 \rangle = 0$ (and thus also $0 \perp v$)
3. $v \perp w$ implies $v \perp aw$, for each $a \in \mathbb{R}$
4. $v \perp w_1$ and $\cdots$ and $v \perp w_n$ implies $v \perp (w_1 + \cdots + w_n)$

Proof: By simply using the properties of inner product.

1. $v \perp w \implies \langle v, w \rangle = 0 \implies \langle w, v \rangle = \langle v, w \rangle = 0 \implies w \perp v$
2. $0 = 0 \cdot \langle v, v \rangle = \langle 0v, v \rangle = \langle 0, v \rangle = \langle v, 0 \rangle$, so $v \perp 0$
3. if $\langle v, w \rangle = 0$, then: $\langle v, aw \rangle = \langle aw, v \rangle = a\langle w, v \rangle = 0$
4. We do the proof for $n = 2$, so assume $v \perp w_1$ and $v \perp w_2$. Then:

$$\langle v, w_1 + w_2 \rangle = \langle w_1 + w_2, v \rangle = \langle w_1, v \rangle + \langle w_2, v \rangle = 0 + 0 = 0.$$
Theorem

For orthogonal vectors $v, w$,

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2$$

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

$$\|v - w\|^2 = \langle v - w, v - w \rangle$$

$$= \langle v, v - w \rangle + \langle -w, v - w \rangle$$

$$= \langle v, v \rangle - \langle v, w \rangle - \langle w, v \rangle + \langle w, w \rangle$$

$$= \langle v, v \rangle - 0 - 0 + \langle w, w \rangle$$

$$= \|v\|^2 + \|w\|^2$$
Lemma

Call a set \{v_1, \ldots, 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 $\implies$ Independence): assume $\{v_1, \ldots, v_n\}$ is orthogonal and $a_1 v_1 + \cdots + a_n v_n = 0$. Then for each $i \leq n$:

\[
0 = \langle 0, v_i \rangle = \langle a_1 v_1 + \cdots + a_n v_n, v_i \rangle = \langle a_1 v_1, v_i \rangle + \cdots + \langle a_n v_n, v_i \rangle = a_1 \langle v_1, v_i \rangle + \cdots + a_n \langle v_n, v_i \rangle = a_i \langle v_i, v_i \rangle \quad \text{since} \quad \langle v_j, v_i \rangle = 0 \text{ for } j \neq i
\]

But since $v_i \neq 0$ we have $\langle v_i, v_i \rangle \neq 0$, and thus $a_i = 0$. This holds for each $i$, so $a_1 = \cdots = a_n = 0$, and we have proven independence.
Consider in $\mathbb{R}^2$ a vector $v = (x_1, x_2)$ that we wish to project onto a line given by another vector $w = (y_1, y_2)$, as in:

\[ w = (y_1, y_2) \]
\[ v = (x_1, x_2) \]

This projection of $v$ on the (line through) $w$ is:

- a multiple $pw$ of $w$, for some $p \in \mathbb{R}$, such that:
- the distance $d(pw, v)$ is minimal.
Projecting a vector on a line, part II

• The distance between \( pw = (py_1, py_2) \) and \( v = (x_1, x_2) \) is:

\[
d(pw, v) = \|pw - v\| = \sqrt{(py_1 - x_1)^2 + (py_2 - x_2)^2}
\]

• Thus, we need to find a \( p \in \mathbb{R} \) for which
  
  - \((py_1 - x_1)^2 + (py_2 - x_2)^2\) is minimal
  - \(\ldots\) where \(x_i\) and \(y_i\) are fixed, and \(p\) is our variable

• We start by simply calculating:

\[
(py_1 - x_1)^2 + (py_2 - x_2)^2
= (py_1)^2 - 2px_1y_1 + x_1^2 + (py_2)^2 - 2px_2y_2 + x_2^2
= (y_1^2 + y_2^2)p^2 - 2(x_1y_1 + x_2y_2)p + x_1^2 + x_2^2
= \|w\|^2p^2 - 2\langle v, w \rangle p + \|v\|^2
\]

• For which \(p\) is this parabola minimal?
Secondary school problem: minimum when derivative is zero

Derivative (in \( p \)) of \( \|w\|^2 p^2 - 2\langle v, w \rangle p + \|v\|^2 \) is:

\[
2\|w\|^2 p - 2\langle v, w \rangle
\]

which is zero when:

\[
p = \frac{\langle v, w \rangle}{\|w\|^2} = \frac{\langle v, w \rangle}{\langle w, w \rangle}
\]

Thus the projection \( v_\parallel \) of \( v \) on \( w \) is the vector

\[
v_\parallel = pw = \frac{\langle v, w \rangle}{\langle w, w \rangle} w
\]

Sometimes we also write \( \pi_w(v) \) for \( v_\parallel \), to make \( w \) explicit.
Next we put $v_{\perp} = v - v_{\parallel}$, so that:

$$\langle v_{\perp}, w \rangle = \langle v - v_{\parallel}, w \rangle = \langle v, w \rangle - \langle v_{\parallel}, w \rangle$$

$$= \langle v, w \rangle - \frac{\langle v, w \rangle}{\langle w, w \rangle} \langle w, w \rangle$$

$$= \langle v, w \rangle - \frac{\langle v, w \rangle}{\langle w, w \rangle} \langle w, w \rangle$$

$$= \langle v, w \rangle - \langle v, w \rangle = 0$$
Theorem

Assume a vector $v \in V$ in a space $V$ with:

- a subspace $W \subseteq V$, where $W$ has orthogonal basis
  $\{w_1, \ldots, w_n\}$

The projection $\pi_W(v)$ of $v$ on the subspace $W$ is then the sum of
projections on the base vectors $w_i$, that is:

$$
\pi_W(v) = \pi_{w_1}(v) + \cdots + \pi_{w_n}(v)
$$

$$
= \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \cdots + \frac{\langle v, w_n \rangle}{\langle w_n, w_n \rangle} w_n
$$

Note: This only works for an orthogonal basis for $W$

- if we have a non-orthogonal basis, we have to make it
  orthogonal first!
  (Using the so called Gram-Schmidt procedure ... later ...)
Projecting a vector on a subspace: simple example

- Suppose in $\mathbb{R}^3$ we wish to project $v = (1, 1, 1)$ on the xy plane (where $z = 0$)
  - the result should be $(1, 1, 0)$; does it come out?
- The obvious orthogonal basis for the xy-plane is $w_1 = (1, 0, 0)$ and $w_2 = (0, 1, 0)$.
  - $\pi_{xy}(v) = \pi_{w_1}(v) + \pi_{w_2}(v) = \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \frac{\langle v, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2$
    - $= \frac{1}{1}(1, 0, 0) + \frac{1}{1}(0, 1, 0) = (1, 1, 0)$
- Another orthogonal bases $u_1 = (1, 2, 0), u_2 = (-2, 1, 0)$ of the xy-plane yields the same outcome:
  - $\pi_{xy}(v) = \pi_{u_1}(v) + \pi_{u_2}(v) = \frac{\langle v, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 + \frac{\langle v, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2$
    - $= \frac{3}{5}(1, 2, 0) + \frac{-1}{5}(-2, 1, 0)$
    - $= (\frac{3}{5}, \frac{6}{5}, 0) + (\frac{-2}{5}, \frac{-1}{5}, 0) = (1, 1, 0)$ ✓
- Check for yourself: this fails for a non-orthogonal basis for the xy-plane (or see LNBS for a similar example)
• Consider a vector $v$ and its projection $\pi_W(v) \in W$ onto a subspace $W$

• This $\pi_W(v) \in W$ is a best approximation of $v$

• More precisely, for an arbitrary $w \in W$, the distance $d(v, w)$ is minimal when $w = \pi_W(v)$.

• This we can verify (see next slide)
Write $v_\perp = v - \pi_W(v)$, then $\langle v_\perp, w \rangle = 0$ for all $w \in W$.

And, for $w \in W$,

$$d(v, w)^2 = \|v - w\|^2 = \| (v - \pi_W(v)) + (\pi_W(v) - w) \|^2$$

$$= \langle (v - \pi_W(v)) + (\pi_W(v) - w), (v - \pi_W(v)) + (\pi_W(v) - w) \rangle$$

$$= \langle v - \pi_W(v), v - \pi_W(v) \rangle + \langle v - \pi_W(v), \pi_W(v) - w \rangle +$$

$$\langle \pi_W(v) - w, v - \pi_W(v) \rangle + \langle \pi_W(v) - w, \pi_W(v) - w \rangle$$

$$= \| v - \pi_W(v) \|^2 + \langle v_\perp, \pi_W(v) - w \rangle -$$

$$\langle \pi_W(v) + w, v_\perp \rangle + \| \pi_W(v) - w \|^2$$

$$= \| v - \pi_W(v) \|^2 + \| \pi_W(v) - w \|^2$$

$$= d(v, \pi_W(v))^2 + d(\pi_W(v), w)^2$$

This second term $d(\pi_W(v), w)^2$ is minimal for $w = \pi_W(v)$. 
• We have seen that, for a set \( \{v_1, \ldots, v_n\} \) of non-zero vectors:

\[
\text{Orthogonality} \quad \iff \quad \text{Independence}
\]

• But we can transform an independent set \( \{v_1, \ldots, v_n\} \) of vectors into an orthogonal set \( \{v'_1, \ldots, v'_n\} \).

• This procedure is called **Gram-Schmidt orthogonalisation**
Gram-Schmidt orthogonalisation, part I

1. Starting point: independent set \( \{v_1, \ldots, v_n\} \) of vectors
2. Take \( v_1' = v_1 \)
3. Take \( v_2' = v_2 - \frac{\langle v_2, v_1' \rangle}{\langle v_1', v_1' \rangle} v_1' \)

[This is \( v_2 \) minus the projection of \( v_2 \) onto \( v_1' \)]

Then:
\[
\langle v_2', v_1' \rangle = \langle v_2 - \frac{\langle v_2, v_1' \rangle}{\langle v_1', v_1' \rangle} v_1', v_1' \rangle
= \langle v_2, v_1' \rangle - \frac{\langle v_2, v_1' \rangle}{\langle v_1', v_1' \rangle} \langle v_1', v_1' \rangle
= \langle v_2, v_1' \rangle - \frac{\langle v_2, v_1' \rangle}{\langle v_1', v_1' \rangle} \langle v_1', v_1' \rangle
= \langle v_2, v_1' \rangle - \langle v_2, v_1' \rangle
= 0
\]

4. \( \ldots \)
Put $v'_i = v_i - \frac{\langle v_i, v'_1 \rangle}{\langle v'_1, v'_1 \rangle} v'_1 - \cdots - \frac{\langle v_i, v'_{i-1} \rangle}{\langle v'_{i-1}, v'_{i-1} \rangle} v'_{i-1}$.

By essentially the same reasoning as before one shows:

$\langle v'_i, v'_j \rangle = 0,$ for all $j < i$.

Result: orthogonal set $\{v'_1, \ldots, v'_n\}$. 
Gram-Schmidt orthogonalisation, illustration

- Take in $\mathbb{R}^4$, $v_1 = (0, 1, 2, 1)$, $v_2 = (0, 1, 3, 1)$, $v_3 = (1, 1, 1, 0)$
- $v'_1 = v_1 = (0, 1, 2, 1)$; then $\langle v'_1, v'_1 \rangle = 1 \cdot 1 + 2 \cdot 2 + 1 \cdot 1 = 6$.
- $v'_2 = v_2 - \frac{\langle v_2, v'_1 \rangle}{\langle v'_1, v'_1 \rangle} v'_1$
  
  \[= (0, 1, 3, 1) - \frac{1 \cdot 1 + 3 \cdot 2 + 1 \cdot 1}{6} (0, 1, 2, 1)\]

  \[= (0, 1, 3, 1) - \frac{8}{6} (0, 1, 2, 1) = (0, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{3})\]

  We prefer to take: $v'_2 = (0, -1, 1, -1)$; then $\langle v'_2, v'_2 \rangle = 3$.
- $v'_3 = v_3 - \frac{\langle v_3, v'_1 \rangle}{\langle v'_1, v'_1 \rangle} v'_1 - \frac{\langle v_3, v'_2 \rangle}{\langle v'_2, v'_2 \rangle} v'_2$
  
  \[= \cdots = (1, \frac{1}{2}, 0, -\frac{1}{2})\]

  We can change it into $v'_3 = (2, 1, 0, -1)$, for convenience.
Orthogonal and orthonormal bases

**Definition**

A basis $B = \{v_1, \ldots, v_n\}$ of a vector space with an inner product is called:

1. **orthogonal** if $B$ is an orthogonal set: $\langle v_i, v_j \rangle = 0$ if $i \neq j$
2. **orthonormal** if it is orthogonal and $\|v_i\| = 1$, for each $i$

**Example**

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

By Gram-Schmidt each basis can be made orthogonal (first), and then orthonormal by replacing $v_i$ by $\frac{1}{\|v_i\|} v_i$. 
The aim of a search engine is to transform a query (question) into an ordered list of documents, with the top one being “most relevant”

Assume $Q$ is the query and $D_1, \ldots, D_n$ the (very long) list of available documents, from which an ordered sublist must be returned (in response to $Q$)

- the query itself is also seen as a document

Idea: use a “metric”: define a suitable distance between documents and return the sublist $D_{i_1}, D_{i_2}, \ldots$ with

$$d(Q, D_{i_1}) \leq d(Q, D_{i_2}) \leq \cdots$$

minimal
Search engines in general, part II

- How to define such a distance between documents — or an inner product giving rise to a distance?
- Google uses PageRank, a link analysis technique
  - a hyperlink to a page counts as a vote of support
  - the PageRank of a page is defined recursively, and is higher if pages with a high PageRank refer to it
- Here we sketch an alternative technique using vectors of word occurrences
- A course in Information Retrieval gives a more systematic account
Comparing occurrence vectors, part I

- Assume we have $N$ terms $t_1, \ldots, t_N$ whose occurrence is relevant.
- For each document $D$ define a vector $\text{occs}(D) \in \mathbb{R}^N$ as:
  \[
  \text{occs}(D) = (\text{occs}_1(D), \ldots, \text{occs}_N(D)),
  \]
  where
  \[
  \text{occs}_i(D) = \text{the number of occurrences of term } t_i \text{ in } D.
  \]
- One way of comparing a query $Q$ and a document $D$ is to look at the angle between vectors $\text{occs}(Q)$ and $\text{occs}(D)$ in the space $\mathbb{R}^N$.
  - Recall the formula:
    \[
    \cos(\gamma) = \frac{\langle \text{occs}(Q), \text{occs}(D) \rangle}{\|\text{occs}(Q)\| \|\text{occs}(D)\|}
    \]
• Not all of the terms $t_i$ are equally relevant
• A more refined approach multiplies $\text{occs}_i(D)$ with the (global) relevance $r_i \in \mathbb{R}$ of the term $t_i$
• One can take $r_i$ for instance as:
\[
    r_i = \log \left( \frac{|D|}{|\{D' \in D \mid t_i \in D'\}|} \right)
\]
where $D$ is the set of all documents.
• We now consider the weight vector $w(D) = (w_1(D), \ldots, w_N(D))$ where
\[
    w_i(D) = r_i \cdot \text{occs}_i(D).
\]
• The angles between weight vectors are computed in the standard way via:
\[
    \cos(\gamma) = \frac{\langle w(Q), w(D) \rangle}{\|w(Q)\| \|w(D)\|}
\]
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 calculation rules
  - they are the focus in this course
  - and are extremely useful in many other settings
About the exam, part I

- Closed book; no calculator / smart phone / ⋯ 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 reproducible literally
    - but help you to understand questions
Calculation rules (or formulas) must be known by heart for:

1. solving (non)homogeneous equations, echelon normal form
2. linearity, independence, matrix application, kernel & image
3. matrix multiplication & inverse, change-of-basis matrices
4. eigenvalues, eigenvectors and determinants (including abc-formula)
5. inner products, distance, length, angle, orthogonality, projection, Gram-Schmidt orthogonalisation
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 2015)

- Exam: Wednesday, April 1, 8:30–11:30 in LIN2.
- **Question time:** Tuesday March 31, 15:45-16.30 in GN7, about course material, last exercises (assignment 8) & exam of spring 2014 (see the webpage).
- Assignment 8: now on the web. Hand in: 30/3 **before 13:00**
- How we compute the final grade \( g \) for the course
  - Your exam grade \( e \), which should be \( \geq 5 \),
  - Your average assignment grade \( a \), which should be \( \geq 5 \).
  - We compute \( e + \frac{a}{10} \) and round it off to a half (but no 5.5).
Students who do the exam for the third (or more) time:

- You should register 1 week before the exam.
- There will be a second chance (hertentamen). Friday July 3, 8:30–11:30, HG00.062
- If you need a 3d time registration see me way in advance in Mercator 1, 00.05 
  (assuming you have participated actively)
- Bring your own filled in form! 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!**