Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Operations Research
Department of Mathematics

Robert Weismantel

Week 8: Orthogonal vectors, orthogonal complements of subspaces and projections

Orthogonality of vectors and subspaces

The target

Orthogonality is a key concept that allows us to decompose a space into two subspaces, understand systems of linear equations, and allows us to define a pseudoinverse.

Definition

Vectors $v, w \in \mathbb{R}^n$ are orthogonal/ perpendicular (see Def. 1.15) if

$$v^T w = \sum_{i=1}^{n} v_i w_i = 0.$$

Subspaces V and W are orthogonal if for all $v \in V$ and $w \in W$, the vectors v and w are orthogonal.

Lemma

Let $v_1, \ldots, v_k$ and $w_1, \ldots, w_l$ be bases of subspace V and W. V and W are orthogonal if and only if v_i and w_j are orthogonal for all i and j.

Preliminaries

Proof of the first lemma

- Suppose V and W are orthogonal. Since $v_i \in V$ for all i and $w_j \in W$ for all j, we have
$$v_i^T w_j = 0 \text{ for all } i, j.$$
- Conversely, assume that $v_i^T w_j = 0$ for all i and j.
- Let $v = \sum_{i=1}^k \lambda_i v_i \in V$ and $w = \sum_{j=1}^l \mu_j w_j \in W$.

$$v^T w = \sum_{i=1}^k \lambda_i v_i^T w = \sum_{i=1}^k \lambda_i v_i^T \sum_{j=1}^l \mu_j w_j = \sum_{i=1}^k \sum_{j=1}^l \mu_j \lambda_i v_i^T w_j = 0.$$

Lemma

Let V and W be two orthogonal subspaces of $\mathbb{R}^n$. Let $v_1, \ldots, v_k$ be a basis of subspace V. Let $w_1, \ldots, w_l$ be a basis of subspace W. The set of vectors $\{v_1, \ldots, v_k, w_1, \ldots, w_l\}$ are linearly independent.

Further preliminaries

Proof of the second lemma

- Consider the linear combination

$$(*) \sum_{i=1}^{k} \lambda_i v_i + \sum_{j=1}^{l} \mu_j w_j = 0.$$

  We want to show $\lambda_i = 0$ for all i and $\mu_j = 0$ for all j.

- Let $v = \sum_{i=1}^{k} \lambda_i v_i$. $(*)$ is equivalent to $v = -\sum_{j=1}^{l} \mu_j w_j$. We obtain

$$v^T v = -\sum_{j=1}^{l} \mu_j v^T w_j = 0.$$

- Hence, $v = 0$. This implies $\lambda_i = 0$ for all i ($v_1, \ldots, v_k$ is a basis of V).
- Accordingly, one shows that $\mu_j = 0$ for all j by considering $w = \sum_{j=1}^{l} \mu_j w_j$ and noticing that $w^T w = 0$.

The orthogonal complement of a subspace I

Corollary

Let V and W be orthogonal subspaces. Then $V \cap W = \{0\}$. Moreover,

$$\dim(V+W) = \dim(\{v+w \mid v \in V,\ w \in W\}) = \dim(V) + \dim(W) \leq n.$$

Definition

Let V be a subspace of $\mathbb{R}^n$. We define the orthogonal complement of V as

$$V^{\perp} = \{w \in \mathbb{R}^n \mid w^T v = 0 \text{ for all } v \in V\}.$$

$V^{\perp}$ is a subspace of $\mathbb{R}^n$!

Theorem

Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Then $N(A) = C(A^T)^{\perp} = R(A)^{\perp}$.

Proof of the Theorem

Proof that $N(A) \subseteq C(A^T)^\perp$.

Let $x \in N(A)$. Take any $b \in C(A^T) = R(A)$, i.e., $b = A^T y$ for some $y \in \mathbb{R}^m$. Then

$$b^T x = y^T A x = y^T 0 = 0.$$

Hence, $x \in C(A^T)^\perp$.

Proof that $C(A^T)^\perp \subseteq N(A)$.

Let $x \in C(A^T)^\perp$. By definition, $b^T x = 0$ for all $b \in C(A^T)$. Define y as the following specific vector: $y := Ax \in \mathbb{R}^m$.
Then $b := A^T y \in C(A^T)$ and hence, $x^T b = 0$. We obtain

$$0 = x^T b = x^T A^T y = x^T A^T A x = \|Ax\|^2 \iff x \in N(A).$$

Recall from Part 1:

If $r = \dim(R(A)) = \dim(C(A^T))$, then $n - r = \dim(N(A))$.

The orthogonal complement of a subspace II

Theorem

Let V, W be orthogonal subspaces of $\mathbb{R}^n$. The statements are equivalent.

(i) $W = V^\perp$.

(ii) $\dim(V) + \dim(W) = n$.

(iii) *Every $u \in \mathbb{R}^n$ can be written as $u = v + w$ with unique $v \in V$, $w \in W$.*

Recall for the proof

Let $v_1, \dots, v_k$ be a basis of V and $w_1, \dots, w_l$ a basis of W. V and W are orthogonal if and only if $v_i^T w_j = 0$ for all $i \in \{1, \dots, k\}$, $j \in \{1, \dots, l\}$.

(i) implies (ii):

Define $A \in \mathbb{R}^{k \times n}$ to be the matrix with row vectors $v_1, \dots, v_k$. Then $V = R(A) = C(A^T)$. Moreover, $W = V^\perp = N(A)$ from the previous theorem. From the remark one slide before:

$$\dim(V) = k \text{ and hence, } \dim(W) = n - k.$$

Proof continued

(ii) implies (iii):

- The vectors in the set $\{v_1,\ldots,v_k,w_1,\ldots,w_l\}$ are linearly independent.
- Since by assumption $l = n-k$, this set is a basis of $\mathbb{R}^n$. Hence,

$$\text{for all } u \in \mathbb{R}^n, \quad u = \sum_{i=1}^{k} \lambda_i v_i + \sum_{j=1}^{l} \mu_j w_j, \text{ where } \lambda_1, \lambda_k, \mu_1, \ldots, \mu_l \in \mathbb{R}.$$

- Define the unique vectors $v := \sum_{i=1}^{k} \lambda_i v_i$, $w := \sum_{j=1}^{l} \mu_j w_j$.

(iii) implies (i): We need to show that $W = V^\perp$.

- $W \subseteq V^\perp$ since W is orthogonal to V.
- For the reverse inclusion, let $u \in V^\perp \subseteq \mathbb{R}^n$. From (iii) $u = v + w$ where $v \in V$ and $w \in W$. Then

$$0 = u^T v = v^T v + v^T w = v^T v = \|v\|^2 \Rightarrow v = 0 \Rightarrow u = w \in W.$$

Decomposition of $\mathbb{R}^n$

Lemma

Let V be a subspace of $\mathbb{R}^n$. Then $V = (V^\perp)^\perp$.

Proof. Let $v_1, \ldots, v_k$ be a basis of V and $w_1, \ldots, w_l$ a basis of $V^\perp$.

- $l = n - k$. Moreover, $v_i^T w_j = 0$ for all i and j and hence,

$$(V^\perp)^\perp = \{x \in \mathbb{R}^n \mid x^T w_j = 0 \text{ for all } j = 1, \ldots, n-k\}.$$

- Since $v_i^T w_j = 0$ for all $j = 1, \ldots, n-k$ we obtain that $V \subseteq (V^\perp)^\perp$. From the Theorem before, $\dim((V^\perp)^\perp) = n - (n-k) = k$.
- Since $\{v_1, \ldots, v_k\} \subseteq V \subseteq (V^\perp)^\perp$ are linearly independent, they are a basis of $(V^\perp)^\perp$. Hence $V = (V^\perp)^\perp$.

Corollary

For a subspace V of $\mathbb{R}^n$, $\mathbb{R}^n = V + V^\perp = \{v + w \mid v \in V,\ w \in V^\perp\}$.

The set of all solutions to a system of linear equations

Corollary

For $A \in \mathbb{R}^{m \times n}$, $N(A) = C(A^T)^\perp$ and $C(A^T) = N(A)^\perp$.

To refine our understanding,

Let $A \in \mathbb{R}^{m \times n}$. There are two important subspaces associated with A:

$$\begin{aligned} N(A) &= \{x \in \mathbb{R}^n \mid Ax = 0\} \\ R(A) &= C(A^T) = \{A^T y \mid y \in \mathbb{R}^m\} = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m \text{ such that } x = A^T y\}. \end{aligned}$$

$N(A)$ is the orthogonal complement of $R(A)$ and $R(A)$ the orthogonal complement of $N(A)$. Hence

$\forall x \in \mathbb{R}^n$ there exist $x_0 \in N(A)$ and $x_1 \in R(A)$ such that $x = x_0 + x_1$ and $x_1^T x_0 = 0$.

Theorem

$$\{x \in \mathbb{R}^n \mid Ax = b\} = x_1 + N(A) \textit{ where } x_1 \in R(A) \textit{ such that } Ax_1 = b.$$

A link between the nullspaces of A and $A^T A$

Lemma

Let $A \in \mathbb{R}^{m \times n}$. Then $N(A) = N(A^T A)$ and $C(A^T) = C(A^T A)$.

Proof.

- If $x \in N(A)$ then $Ax = 0$ and so $A^\top Ax = 0$, thus $x \in N(A^\top A)$.
- If $x \in N(A^\top A)$ then $A^\top Ax = 0$. This implies that

$$x^\top A^\top Ax = x^\top 0 = 0.$$

  This gives

- $$0 = x^\top A^\top Ax = (Ax)^\top (Ax) = \|Ax\|^2,$$

  so $Ax = 0$ and so $x \in N(A)$.
- For the second statement we notice

$$C(A^T) = N(A)^\perp = N(A^T A)^\perp = C((A^T A)^T) = C(A^T A).$$

Projections

Definition (Projection of a vector onto a subspace)

The projection of a vector $b \in \mathbb{R}^m$ on a subspace S (of $\mathbb{R}^m$) is the point in S that is closest to b. In other words

$$\mathrm{proj}_S(b) = \operatorname*{argmin}_{p \in S} \|b - p\|. \tag{1}$$

Sanity check

This is only a proper definition if the minimum exists and is unique.

The one-dimensional case

Let S be the subspace corresponding to the line that goes through the vector $a \in \mathbb{R}^m \setminus \{0\}$, i.e. $S = \{\lambda a \mid \lambda \in \mathbb{R}\} = C(a)$. By drawing a two dimensional example one can see that the projection p is the vector in the subspace S such that the “error vector” $e = b - p$ is perpendicular to a (i.e. $b - p \perp a$).

The one dimensional case

Lemma

Let $a \in \mathbb{R}^m \setminus \{0\}$. The projection of $b \in \mathbb{R}^m$ on $S = \{\lambda a \mid \lambda \in \mathbb{R}\} = C(a)$ is

$$\text{proj}_S(b) = \frac{aa^T}{a^T a} b.$$

Proof. Let $p \in S$, $p = \lambda a$ for $\lambda \in \mathbb{R}$.

$$\|b-p\|^2 = (b-p)^T(b-p) = b^T b - 2b^T p + p^T p = \|b\|^2 - 2\lambda b^T a + \lambda^2 \|a\|^2 = g(\lambda).$$

g is a convex, quadratic function in one variable λ.

The minimizer is obtained at the point λ^* where the derivative vanishes.

$$g'(\lambda) = -2b^T a + 2\lambda \|a\|^2 = 0 \iff \lambda^* = \frac{b^T a}{a^T a}.$$

$$\text{Hence, } \text{proj}_S(b) = \lambda^* a = a\frac{b^T a}{a^T a} = a\frac{a^T b}{a^T a} = \frac{aa^T}{a^T a} b.$$

About our initial intuition

Our guess was that

the projection p should be the vector in the subspace S such that the "error vector" $e = b - p$ is perpendicular to a, i.e.,

$$(b - \text{proj}_S(b)) \perp a.$$

By substituting what we just computed we get

$$a^T(b - \text{proj}_S(b)) = a^T(b - \frac{aa^T}{a^Ta}b) = a^Tb - a^T(\frac{aa^T}{a^Ta}b) =$$

$$a^Tb - \frac{1}{a^Ta}a^Taa^Tb = a^Tb - a^Tb = 0.$$

A final check

The projection of a vector that is already a multiple of a should be the vector itself. This is indeed true!

The general case

The idea is similiar to the one-dimensional case

Let S be a subspace in $\mathbb{R}^m$ generated by $a_1, \ldots, a_n \in S$, i.e.,

$$S = \text{ span}(a_1, \ldots, a_n) = C(A) = \{Ax \mid x \in \mathbb{R}^n\}$$

where

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}.$$

Lemma

The projection of a vector $b \in \mathbb{R}^m$ to the subspace $S = C(A)$ can be written as

$$\text{proj}_S(b) = A\hat{x}, \textit{ where } \hat{x} \textit{ satisfies the normal equations } A^T A\hat{x} = A^T b.$$

Recall for $m = 1$

$$\text{proj}_S(b) = \lambda^* a = \frac{aa^T}{a^T a} b \iff a^T a \lambda^* a = a^T ba \iff a^T a \lambda^* = a^T b.$$

Proof.

- $b \in \mathbb{R}^m$. Hence $b = p + e$ where $p \in S$ and $e \in S^\perp$, i.e., $p^T e = 0$.
- Consider another point $p' \in S$. Then $p - p' \in S$ and hence, $e^T(p - p') = 0$. This gives

$$\begin{aligned} \|p' - b\|^2 &= \|p' - p + p - b\|^2 = \|p' - p - e\|^2 \\ &= \|p' - p\|^2 + \|e\|^2 \geq \|e\|^2 = \|p - b\|^2. \end{aligned}$$

- We have shown that

$$\text{proj}_S(b) = p = A\hat{x} \in S$$

  where $b = p + e$ with $e \in S^\perp$.
- Since $S = C(A)$,

$$(b - \text{proj}_S(b)) \perp a_i \text{ for all } i = 1, \ldots, n \iff a_i^T(b - \text{proj}_S(b)) = 0 \text{ for all } i.$$

- This is equivalent to saying that

$$A^T(b - \text{proj}_S(b)) = 0 \iff A^T(b - A\hat{x}) = 0 \iff A^T A\hat{x} = A^T b.$$