Basis
Norm

Peter Büchel

HSLU W

March 19, 2021

Basis and Dimension

- Aim: Span a subspace with the *least* numbers of vectors
- Linear independence of vectors: Condition that span is as *large* as possible

Definition

Let U be a linear subspace of $\mathbb{R}^n$.
The vectors $v_1, \ldots, v_k \in U$ are a *basis* of U, if

- they generate the whole subspace U, i.e., $U = \text{span}(v_1, \ldots, v_k)$
- they are linearly independent

Remarks

- Let $v_1, \dots, v_k \in U$ be a basis of U
- Then every vector $v \in U$ can be *uniquely* written as a linear combination

$$v = s_1 v_1 + \cdots + s_k v_k \tag{1}$$

Proof

- Property a) of the definition of basis: Each vector of U *can* be written as a linear combination of $v_1, \ldots, v_k$
- What remains to be shown is the uniqueness
- So let us assume that the vector $v \in U$ has two representations:

$$v = s_1 v_1 + \cdots + s_k v_k \qquad \text{and} \qquad v = t_1 v_1 + \cdots + t_k v_k$$

- Subtraction of these two equations yields

$$0 = v - v = s_1 v_1 + \cdots + s_k v_k - (t_1 v_1 + \cdots + t_k v_k) = (s_1 - t_1) v_1 + \cdots + (s_k - t_k) v_k$$

- Linear combination of the zero vector with the coefficients $s_1 - t_1, \ldots, s_k - t_k$
- Since the $v_1, \ldots, v_n$ are linearly independent, we conclude

$$s_i - t_i = 0 \quad \Rightarrow \quad s_i = t_i \quad \text{for all} \quad 1 \leq i \leq k$$

- So the two representations are in fact identical: Uniqueness

Notation

- Coefficients $s_1, \ldots, s_k$ in (1): *Coordinates* of v with respect the basis $v_1, \ldots, v_k$

- Write the coordinates as a column, the so-called *coordinate vector*:

$$\begin{bmatrix} s_1 \\ \vdots \\ s_k \end{bmatrix} \in \mathbb{R}^k$$

- Instead of considering the elements of U as column vectors with n components, we can also represent them by their coordinate vectors with k components

- Thus, if $k < n$, then we need not store as many numbers

- So, we need less memory and less computational effort

- This can be seen as a kind of *data compression*.

Example

- *Standard basis* (or *canonical basis*) of $\mathbb{R}^n$:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \dots \quad e_{n-1} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} \quad e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

- The coordinates of the vectors $x \in \mathbb{R}^n$ with respect to the standard basis are its component, since

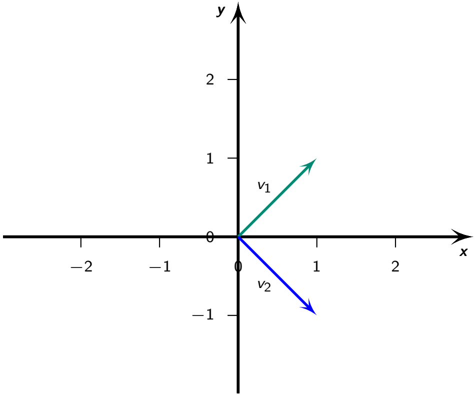
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1 e_1 + \cdots x_n e_n$$

Example

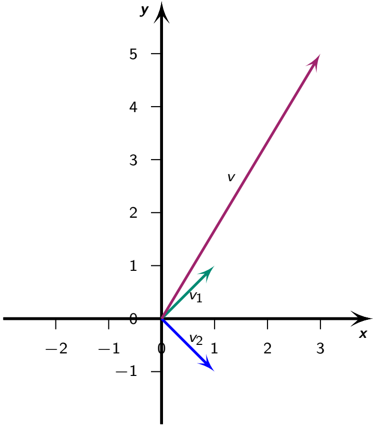
- The following vectors generate the whole space $\mathbb{R}^2$:

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- Sketch:



- Since they are linearly independent, they form a basis of $\mathbb{R}^2$
- Example: Determine coordinates of the vector $v = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ with respect to the basis v_1, v_2
- Sketch:



- Need to find scalars s_1, s_2 so that $v = s_1 v_1 + s_2 v_2$, i.e.,

$$\begin{bmatrix} 3 \\ 5 \end{bmatrix} = s_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + s_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} s_1 + s_2 \\ s_1 - s_2 \end{bmatrix}$$

- Obtain a linear system of equations: Solve with the Gaussian elimination method:

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 1 & -1 & 5 \end{array}\right]_{Z_{12}(-1)} \Leftrightarrow \quad \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -2 & 2 \end{array}\right]$$

- Thus $s_2 = -1$ and so $s_1 = 3 - s_2 = 4$

- Therefore:

$$v = 4v_1 - v_2$$

- Coordinate vector of v:

$$\begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

Dimension

- Each basis of a vector subspace U has the same number of elements
- This number is called the *dimension* of U, written $\dim U$
- Example: Standardbasis shows: $\dim \mathbb{R}^n = n$

- Starting with a vector $v_1 \neq 0$ from a vector subspace U of $\mathbb{R}^n$
- Repeatedly search for additional linearly independent vectors
- At some point, can not find any linear independent vector anymore
- If we have then obtained k linearly independent vectors $v_1, \ldots, v_k$, each additional vector $v \in U$ must be a linear combination of $v_1, \ldots, v_k$
- Therefore, these vectors generate U

- So they form a basis and the dimension of U is k

Dimension

Let $U \subseteq \mathbb{R}^n$ be a subspace and $v_1, \dots, v_k \in U$

- If $v_1, \dots, v_k$ are linearly independent, then $k \leq \dim U$
- If $V = \operatorname{span}(v_1, \dots, v_k)$, then $k \geq \dim U$
- If $k = \dim U$, then we have

$$\begin{aligned} v_1, \dots, v_k \quad \text{is a basis of } U \quad &\iff \quad v_1, \dots, v_k \quad \text{are linearly independent} \\ &\iff \quad U = \operatorname{span}(v_1, \dots, v_k) \end{aligned}$$

- 3rd Statement: Useful to determine a basis of a subspace when we know its dimension.

Example

- Do the vectors

$$b_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad b_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

form a basis of $\mathbb{R}^2$?

- It is almost obvious that none of the vectors b_1, b_2 is a scalar multiple of the other one, so they are linearly independent

- To check this formally, we calculate the determinant

$$\det \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{vmatrix} 1 & -2 \\ 2 & 1 \end{vmatrix} = 5$$

- Since it is different from zero, the vectors are indeed linearly independent

- Furthermore, we know from that $\dim \mathbb{R}^2 = 2$

- 3rd statement of the previous box implies that the vectors b_1, b_2 form a basis of $\mathbb{R}^2$

- How can we find a basis of a vector subspace U of $\mathbb{R}^n$?
- It certainly depends on how the subspace is given
- If it is given as the span of a finite family of vectors, then an algorithm can be given

Algorithm: Finding a basis of $U = \text{span}(v_1, \ldots, v_k) \subseteq \mathbb{R}^n$

- Write down the $(n \times k)$-matrix consisting of the column vectors v_i

$$A = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}$$

- Transform A in a echelon row form by using elementary row transformations.

- Let $j_1, \ldots, j_p$ be the whole set of columns containing pivot elements, then the vectors

$$v_{j_1}, \ldots, v_{j_p} \quad \text{a basis of } U$$

- The above procedure tells us that there is always a subset of the given vectors which is a basis

- We need the previous algorithm only to *find* this subset

- Note that the row operations change the column vectors, so the new ones may not even be contained in U!

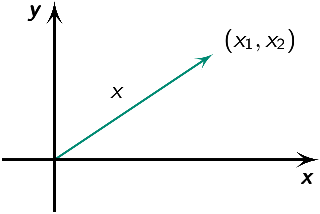
- The procedure only finds the *indices* $j_1, \dots, j_p$, such that the *original* vectors $v_{j_1}, \dots, v_{j_p}$ from a basis of U

Be careful

It is important to take the original vectors and not the ones in the rows echelon form.

Norm, dot product

- Motivation of norm: Consider $\mathbb{R}^2$ and $\mathbb{R}^3$
- Vectors: As arrows with initial point in origin



- The length of a vector x in $\mathbb{R}^2$ and $\mathbb{R}^3$: *Norm* of x
- Denoted by: $\|x\|$

- It follows from the Pythagorean Theorem for $x \in \mathbb{R}^2$:

$$\|x\| = \sqrt{x_1^2 + x_2^2}$$

- For $x \in \mathbb{R}^3$:

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

- Generalisation to $\mathbb{R}^n$ obvious (though no sketches for $n > 3$):

$$\|x\| = \sqrt{x_1^2 + \ldots + x_n^2}$$

- Norm not linear in $\mathbb{R}^n$

Definition of norm

- Definition:

Definition Dot product

For $x, y \in \mathbb{R}^n$, the *dot product* of x and y is defined by

$$x \bullet y = x_1 y_1 + \ldots + x_n y_n$$

- Also $x \cdot y$ or $\langle x, y \rangle$
- $\langle \ldots, \ldots \rangle$ also used for more general *inner product*
- Note: Dot product is a *number*, not a vector

- Dot product in $\mathbb{R}^n$ has the following properties:
  - $x \bullet x \geq 0$
  - $x \cdot x = 0$ if and only if $x = 0$
  - For $y \in \mathbb{R}^n$ fixed, $x \bullet y$ is linear:
    - ★ $(x + z) \bullet y = x \bullet +z \bullet y)$
    - ★ $(\lambda x) \bullet y = \lambda(x \bullet y)$
  - $x \bullet y = y \bullet x$
- Obviously for all $x \in \mathbb{R}^n$:

$$x \bullet x = \|x\|^2$$

Norm

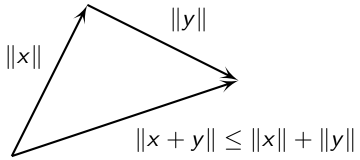
- Norm more general than described above

Definition Norm

A *norm* $\|\cdot\|$ on vector space $\mathbb{R}^n$ is function $\mathbb{R}^n \to \mathbb{R}$ which assigns each vector $x \in \mathbb{R}^n$ a *length* $\|x\| \in \mathbb{R}$ such that for all $\lambda \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$ the following hold:

- $\|\lambda x\| = |\lambda| \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$

- Triangle inequality:

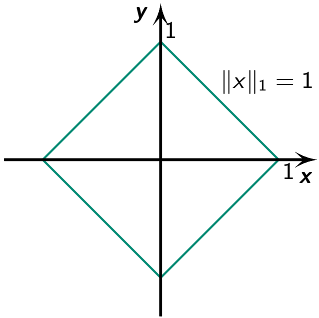


Example: Manhattan Norm

- Manhattan norm on $\mathbb{R}^n$:

$$\|x\|_1 = |x_1| + \ldots + |x_n|$$

- Figure: All vectors $x \in \mathbb{R}^2$ with $\|x\|_1 = 1$:

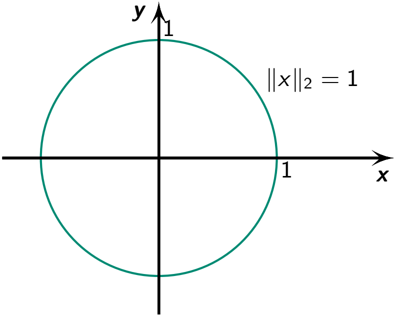


Example: Euclidean Norm

- Already seen:

$$\|x\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$$

- Figure: All vectors $x \in \mathbb{R}^2$ with $\|x\|_1 = 1$:



- From now on: $\|\cdot\| = \|\cdot\|_2$

Example: Hamming distance

- Important in information theory
- Vector in $\mathbb{R}^n$: Components can take only values 0 and 1
- Example in $\mathbb{R}^n$:

$$x = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

- Vector x contains information, that is being sent through a channel (TV, internet)

- Due to "noise": Information doesn't arrive exactly as x, but

$$y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

- Hamming distance: Number of entries, which arrived incorrectly:

$$\|x - y\|_H = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = 1$$

- Number of 1's and -1's in difference vector: Number of incorrect transmitted entries

- $\|x - y\|_H = 5$: All entries were transmitted incorrectly
- $\|x - y\|_H = 0$: All entries were transmitted correctly

Remark: Angles

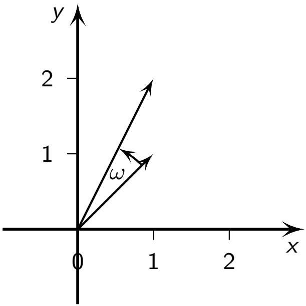
- We can determine the angle between two vectors $x, v \neq 0$ in $\mathbb{R}^2$:

$$\cos\omega = \frac{x \bullet y}{\|x\|\|y\|}$$

- Example: Determine angle between

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

- Sketch:



- For ω:

$$\cos\omega = \frac{x \bullet y}{\|x\|\|y\|} = \frac{1+2}{\sqrt{1^1+1^1}\sqrt{1^2+2^2}} = \frac{3}{\sqrt{10}} = 0.6882472$$

- Hence:

$$\omega = \arccos(0.6882472) = 0.32 \approx 18^\circ$$

Special case: Perpendicular vectors

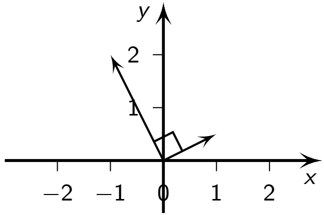
- For $x, y \in \mathbb{R}^2$: Vectors x are y perpendicular if and only of $x \bullet y = 0$

$$x \bullet y = 0 \quad \Longleftrightarrow \quad x \perp y$$

- Example:

$$x = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

- Sketch:



- Dot product:

$$x \bullet y = 1(-1 + 0.5 \cdot 2) = 0$$

Orthogonal vectors

- Definiton:

Definition Orthogonal vectors

Two vectors $u, v \in \mathbb{R}^n$ are called *orthogonal* if $x \bullet y = 0$

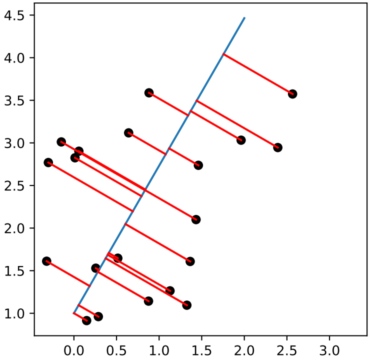
- Orthogonal is just a fancy word for *perpendicular*

Orthogonal projection

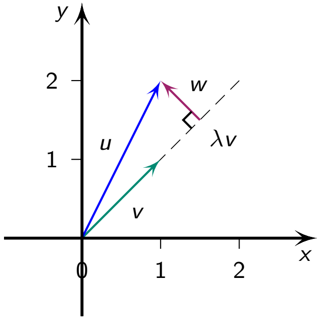
- Important tool in machine learning: Principal component analysis (PCA)

- Orthogonal projection of data point to a subspace of $\mathbb{R}^n$

- Example:



- Sketch:



- We're interested of orthogonal projeciton of u on v

- Hence: Find λ

- Because λv and w orthogonal:

$$(v) \bullet w = 0$$

- But $w = u - \lambda v$
- Hence:

$$v \bullet w = v \bullet (u - \lambda v) = v \bullet u - \lambda v \bullet v = u \bullet v - \lambda \|v\|^2 = 0$$

- Solving for λ:

$$\lambda = \frac{u \bullet v}{\|v\|^2}$$

Example

- Let
$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
- Then:
$$\lambda = \frac{u \bullet v}{\|v\|^2} = \frac{1 \cdot 1 + 1 \cdot 2}{1^2 + 1^2} = \frac{3}{2} = 1.5$$
- Therefore
$$w = u - \lambda v = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$
- We can now write
$$u = \lambda v + w = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$
- Hence, u is the sum of two vectors which are orthognal
- Orthogonal decomposition of u

An orthogonal decomposition

- Suppose $u, v \in \mathbb{R}^n$, with $v \neq 0$
- Set
$$\lambda = \frac{u \bullet v}{\|v\|^2} \qquad \text{and} \qquad w = u - \frac{u \bullet v}{\|v\|^2} v$$
- Then:
$$w \bullet v = 0 \qquad \text{and} \qquad u = \lambda v + w$$