Orthonormal Basis, Gram-Schmidt Process ($\mathbb{R}^n$)

Pascal Pilz

KO Mathematics for AI II

September 2, 2022

Table of Contents

Table of Contents

Definition (Vector space)

By a **vector space** we shall mean a set V on which there are defined two operations, one called "addition" and the other called "multiplication by scalars", such that the following properties hold:

(V_1) $x + y = y + x$ for all $x, y \in V$;

(V_2) $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$;

(V_3) there exists an element $0 \in V$ such that $x + 0 = x$ for every $x \in V$;

(V_4) for every $x \in V$ there exists $-x \in V$ such that $x + (-x) = 0$;

(V_5) $\lambda(x + y) = \lambda x + \lambda y$ for all $x, y \in V$ and all scalars λ;

(V_6) $(\lambda + \mu)x = \lambda x + \mu x$ for all $x \in V$ and all scalars λ, μ;

(V_7) $(\lambda\mu)x = \lambda(\mu x)$ for all $x \in V$ and all scalars λ, μ;

(V_8) $1x = x$ for all $x \in V$.

When the scalars are all real numbers we shall often talk of a **real** vector space; when the scalars are all complex numbers we shall talk of a **complex** vector space.

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 69)

Inner product

Definition (Inner product)

Let V be a vector space over $\mathbb{C}$. By an **inner product** on V we shall mean a mapping $f : V \times V \to \mathbb{C}$, described by $(x, y) \mapsto \langle x|y\rangle$, such that for all $x, x', y \in V$ and for all $\alpha \in \mathbb{C}$, the following identities hold:

$$
\begin{aligned}
&(1) \quad \langle x + x'|y\rangle = \langle x|y\rangle + \langle x'|y\rangle;\\
&(2) \quad \langle \alpha x|y\rangle = \alpha\langle x|y\rangle;\\
&(3) \quad \overline{\langle x|y\rangle} = \langle y|x\rangle \text{ so that in particular } \langle x|x\rangle \in \mathbb{R};\\
&(4) \quad \langle x|x\rangle \geq 0, \text{ with euqality if and only if } x = 0_V.
\end{aligned}
$$

By a **complex inner product space** we mean a vector space V over $\mathbb{C}$ together with an inner product on V. By a **real inner product space** we mean a vector space V over $\mathbb{R}$ together with an inner product on V.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 11)

Example Inner product

Example

On the vector space $\mathbb{R}^n$ of n-tuples of real numbers let

$$\langle (x_1, \ldots, x_n) | (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i y_i .$$

Then it is readily verified that this defines an inner product on $\mathbb{R}^n$, called the **standard inner product** on $\mathbb{R}^n$.
In the cases where $n = 2, 3$ this inner product is often called the **dot product** or **scalar product**. This terminology is popular when dealing with the geometric applications of vectors. Indeed, several of the reults that we shall establish will generalise familiar results in euclidian geometry of two and three dimesnions.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 12, Example 1.1)

Span

Definition (Span)

Let V be a vector space over a field F and let S be a non-empty subset of V. Then we say $v \in V$ is a **linear combination of elements of S** if there exists $x_1, \ldots, x_n \in S$ and $\lambda_1, \ldots, \lambda_n \in F$ such that

$$v = \lambda_1 x_1 + \cdots + \lambda_n x_n = \sum_{i=1}^{n} \lambda_i x_i.$$

It is clear that if $v = \sum_{i=1}^{n} \lambda_i x_i$ and $w = \sum_{i=1}^{n} \mu_i y_i$ are linear combination of elements of S then so is $v + w$; moreover, so is λv for every $\lambda \in F$. Thus the set of linear combinations of elements of S is a subspace of V. We call this the **subspace spanned by S** and denote it by Span S.

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 75-76)

Example Span I

Example

Consider the subset $S_1 = \{(1,0),(0,1)\}$ of the cartesian plane $\mathbb{R}^2$. For every $(x,y) \in \mathbb{R}^2$ we have

$$(x,y) = (x,0) + (0,y) = x(1,0) + y(0,1),$$

so that every element of $\mathbb{R}^2$ is a linear combination of elements of S_1. Thus S_1 is a spanning set of $\mathbb{R}^2$.

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 76, Example 5.14)

Example Span II

Example

Consider the subset $S_2 = \{(1,2), (-1,-1/2)\}$ of the cartesian plane $\mathbb{R}^2$. For every $(x,y) \in \mathbb{R}^2$ we have

$$(x,y) = \frac{1}{3}(2y-x) \cdot (1,2) + \frac{2}{3}(y-2x) \cdot (-1,-1/2),$$

so that every element of $\mathbb{R}^2$ is a linear combination of elements of S_2. Thus S_2 is a spanning set of $\mathbb{R}^2$.

Linearly independent

Definition (Linearly independent)

Let S be a non-empty subset of a vector space V over a field F. Then S is said to be **linearly independent** if the only way of expressing 0_V as a linear combination of elements of S is the trivial way (in which all scalars are 0_F. Equivalently, S is linearly independent if, for any given $x_1, \ldots, x_n \in S$, we have

$$\lambda_1 x_1 + \ldots \lambda_n x_n = 0_V \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_n = 0_F.$$

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 78)

Linear Independence, Basis

Theorem (Linear independence, linear combination)

Let V be a vector space over a field F. If S is a subset of V that contains at least two elements then the following statements are equivalent:

- *S is linearly dependent;*
- *at least one element of S can be expressed as a linear combination of the other elements of S.*

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 78, Theorem 5.5)

Linear Independence, Basis

Theorem (Linear independence, linear combination)

Let V be a vector space over a field F. If S is a subset of V that contains at least two elements then the following statements are equivalent:

- *S is linearly dependent;*
- *at least one element of S can be expressed as a linear combination of the other elements of S.*

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 78, Theorem 5.5)

Definition (Basis)

A **basis** of a vector space V is a **linearly independent** subset of V that **spans** V.

(Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 80)

Orthogonal, Orthonormal, Orthonormal Basis

In the following we mean by $\|x\| = \sqrt{\langle x|x\rangle}$. For more detail see *Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 11*.

Orthogonal, Orthonormal, Orthonormal Basis

In the following we mean by $\|x\| = \sqrt{\langle x|x\rangle}$. For more detail see *Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 11*.

Definition (Orthogonal, orthonormal)

If V is an inner product space then $x, y \in V$ are said to be **orthogonal** if $\langle x|y\rangle = 0$. A non-empty subset S of V is said to be an **orthogonal subset** if every pair of distinct elements of S is orthogonal. An **orthonormal subset** of V is an orthogonal subset S such that $\|x\| = 1$ for every $x \in S$. (Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 17)

Orthogonal, Orthonormal, Orthonormal Basis

In the following we mean by $\|x\| = \sqrt{\langle x|x\rangle}$. For more detail see *Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 11*.

Definition (Orthogonal, orthonormal)

If V is an inner product space then $x, y \in V$ are said to be **orthogonal** if $\langle x|y\rangle = 0$. A non-empty subset S of V is said to be an **orthogonal subset** if every pair of distinct elements of S is orthogonal. An **orthonormal subset** of V is an orthogonal subset S such that $\|x\| = 1$ for every $x \in S$.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 17)

Definition (Orthonormal basis)

By an **orthonormal basis** of a inner product space we mean an orthonormal subset that is a basis.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 17)

Examples Orthonormal Basis

Example (I)

The standard bases of $\mathbb{R}^n$ and of $\mathbb{C}^n$ are orthonormal bases.
I.e., for $\mathbb{R}^n$ we have $e_1, \ldots, e_n$ with $e_i = (\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0)$.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 16, Example 1.10)

Examples Orthonormal Basis

Example (I)

The standard bases of $\mathbb{R}^n$ and of $\mathbb{C}^n$ are orthonormal bases.
I.e., for $\mathbb{R}^n$ we have $e_1, \dots, e_n$ with $e_i = (\underbrace{0, \dots, 0}_{i-1}, 1, 0, \dots, 0)$.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 16, Example 1.10)

Example (II)

In $Mat_{n\times n}\mathbb{C}$ with $\langle A|B\rangle = \sum_{i=1}^{n} c_{ii}$ for $[c_{ij}]_{n\times n} = B * A$ an orthonormal basis is $\{E_{p,q}; p, q = 1, \cdots, n\}$ where $E_{p,q}$ has a 1 in the (p, q)-th position and 0 elsewhere.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 18, Example 1.11)

Table of Contents

Gram-Schmidt orthonormalization process, what is it good for?

We **take a set of vectors** X **that spans a vector space** V, i.e., is a basis of V, and **return another set of vector** Y **that spans the same space**. The set Y is orthonormal.
Consider for example the vector space $\mathbb{R}^2$, the cartesian plane. Vectors can here commonly be represented as arrows, starting at the origin $(0, 0)$ and ending up at the point they describe (x_1, x_2). The commonplace basis for this space are the two unit vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$.
Nevertheless this space can actually be spanned by any two linearly independent vectors. That is, any two vectors that do not point in the same direction can act as a basis for this space.

Gram-Schmidt orthonormalization process

Theorem (Gram-Schmidt orthonormalization process)

Let V be an inner product space and for every non-zero $x \in V$ let $x^* = x/\|x\|$. If $\{x_1, \dots, x_k\}$ is a linearly independent subset of V, define recursively

$$
\begin{aligned}
y_1 &= x_1^*;\\
y_2 &= (x_2 - \langle x_2 | y_1 \rangle y_1)^*;\\
y_3 &= (x_3 - \langle x_3 | y_2 \rangle y_2 - \langle x_3 | y_1 \rangle y_1)^*;\\
&\ \vdots\\
y_k &= \left(x_k - \sum_{i=1}^{k-1} \langle x_k | y_i \rangle y_i \right)^*.
\end{aligned}
$$

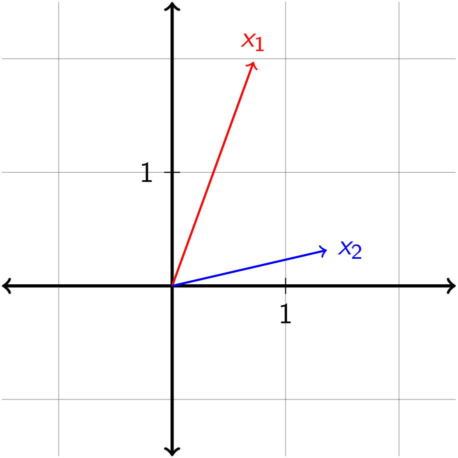
Then $\{y_1, \dots, k_k\}$ is orthonormal and spans the same subspace as $\{x_1, \dots, x_k\}$.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 18, Th. 1.4)

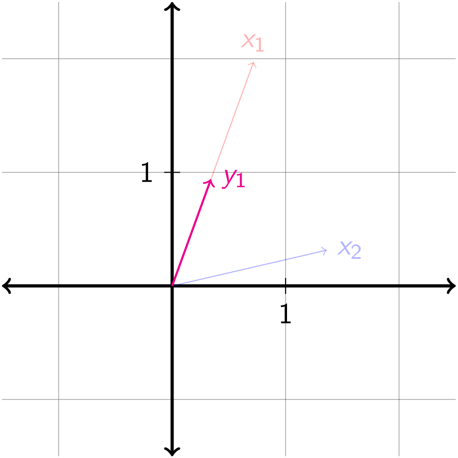
Gram-Schmidt process in words

In words, using the intuitive notion of arrows on vectors in $\mathbb{R}^2$:
A (random) first vector x_1 gets normalized to length 1, the normalized vector is y_1. A next vector x_2 gets chosen, from this vector, we subtract the part of its direction pointing towards the first vector by getting the inner product $\langle y_1 | x_2 \rangle$, which so-to-say gives the length of the projection of x_2 onto y_1, now we multiply this scalar we got from the inner product with the direction of y_1 and subtract this value form x_2. This leaves us with only the part of the direction pointing perpendicular to the first vector. This part now gets normalized to length 1 and becomes y_2.
The process continues like this, taking into consideration all previously established vectors.

Gram-Schmidt process visualization

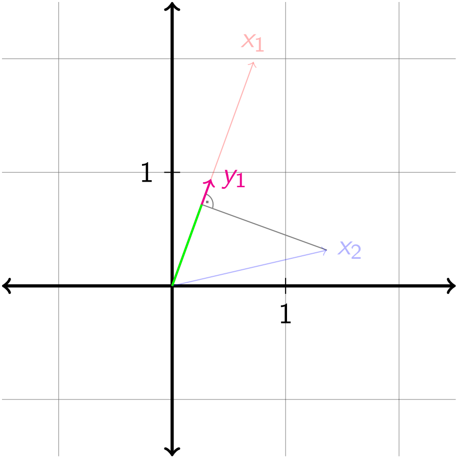


Gram-Schmidt process visualization



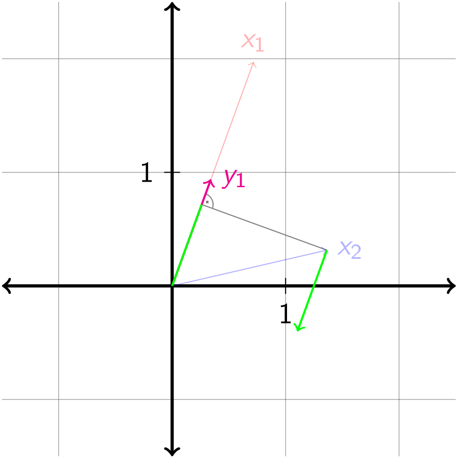
1. Normalize x_1 to get y_1

Gram-Schmidt process visualization



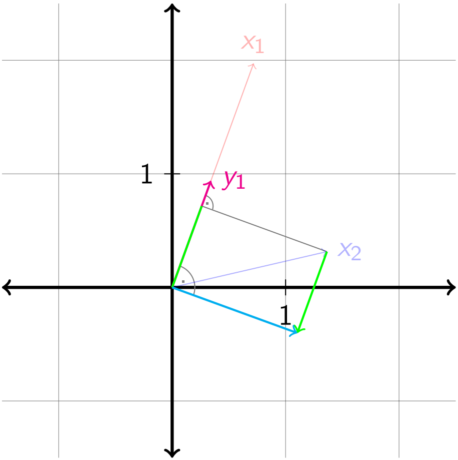
1. Normalize x_1 to get y_1
2. Green is x_2 projected onto y_1, aka $\langle x_2 | y_1 \rangle y_1$

Gram-Schmidt process visualization



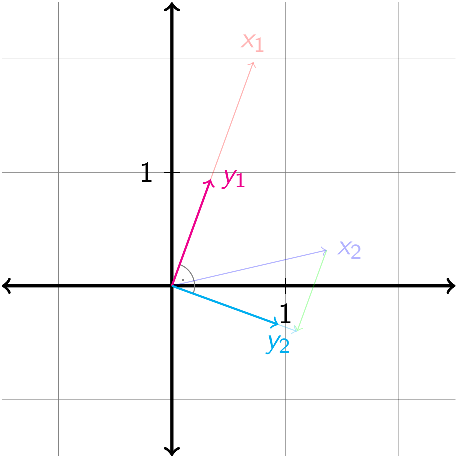
1. Normalize x_1 to get y_1
2. Green is x_2 projected onto y_1, aka $\langle x_2 | y_1 \rangle y_1$
3. Subtract green from x_2

Gram-Schmidt process visualization



1. Normalize x_1 to get y_1
2. Green is x_2 projected onto y_1, aka $\langle x_2 | y_1 \rangle y_1$
3. Subtract green from x_2
4. What you get is a vector perpendicular to y_1. Now normalize this vector to length 1

Gram-Schmidt process visualization



1. Normalize x_1 to get y_1
2. Green is x_2 projected onto y_1, aka $\langle x_2 | y_1 \rangle y_1$
3. Subtract green from x_2
4. What you get is a vector perpendicular to y_1. Now normalize this vector to length 1
5. The result is y_2. y_2 is perpendicular to y_1 and has a length of 1, i.e., $\langle y_1 | y_2 \rangle = 0$ and $\|y_1\| = \|y_2\| = 1$. Whence $\{y_1, y_2\}$ is orthonormal

Example Gram-Schmidt Process

Example

Consider the basis x_1, x_2 of $\mathbb{R}^2$ with $x_1 = (1,2), x_2 = (-1,-1/2)$, the inner product defined as $\langle x|y\rangle = \sum_{i=1}^{n} x_i y_i$ and the norm defined as $\|x\| = \sqrt{\langle x|x\rangle}$.
Applying the Gram-Schmidt process yields

$$y_1 = x_1^* = x_1/\|x_1\| = (1,2)/\sqrt{\langle (1,2)|(1,2)\rangle} = (1/\sqrt{5}, 2/\sqrt{5}) \tag{1}$$

$$y_2 = \left[(-1,-1/2) - \langle (-1,-1/2)|(1/\sqrt{5},2/\sqrt{5})\rangle (1/\sqrt{5},2/\sqrt{5})\right]^* \tag{2}$$

$$= (-6/10, 3/10)^* = (-6/10, 3/10)/(3\sqrt{5}/10) = (-2/\sqrt{5}, 1/\sqrt{5})$$

An interesting result of this:

An interesting result of this:

Corollary

If V is a finite-dimensional inner product space then V has an orthonormal basis.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 19)

An interesting result of this:

Corollary

If V is a finite-dimensional inner product space then V has an orthonormal basis.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 19)

Proof.

Apply the Gram-Schmidt process to a basis of V.

(Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, p. 19) $\square$

Table of Contents

Code

```
def gram_schmidt_process_numpy(s: list) -> list:
    """
    Generates am orthonormal base spanning the same space as the input-base.

            Parameters:
                    s (list): the base spanning the desired space

            Returns:
                base (list): the generated orthonormal base
    """
    # A little helper function to keep it cleaner
    def normalize(vector: np.ndarray) -> np.ndarray:
        return vector / norm(vector)

    s = np.array(s)

    y_k = list()
    for x_k in s:
        element = 0
        for y_i in y_k or [0]:
            element += dot(x_k, y_i) * y_i
        y_k.append(normalize(x_k - element))
    return [list(y_i) for y_i in y_k]
```

Figure: code utilizing numpy

The most difficult part for me was input handling. Numpy provides such vast functionality that implementing the actual algorithm was rather easy.
Implementing the manual version was trickier, as it was necessary to define the various functionalities manually. It might be easier to work with classes, but I decided to not use OOP at all. The manual implementation of the operations allows us to use the same code for the Gram-Schmidt process with different definitions of operations such as inner product.
For more detail see the python files accompanying this presentation.

Recommended Literature

- Basic Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, Chapter 5
- Further Linear Algebra, Blyth, T.S. and Robertson, E.F., 2002, Chapter 1
- Math for AI 1-3 manuscript from WS19-WS20, Ullrich M., 2021, Chapter 11
- Essence of linear algebra (video series), Sanderson G. (3blue1brown), 2016, accessed 15.08.2022, `https://www.3blue1brown.com/topics/linear-algebra`, `https://youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab`

Thank you for your attention!

Presentation made by Pascal Pilz, 12111234, pasc.pilz@gmail.com, for the course "KO Mathematics for AI II", 324.813, by Jan-Michael Holzinger, at Johannes Kepler University Linz.

For the code implementation see
`https://github.com/pilzpascal/KO_Mathematics_for_AI_II`