Linear Algebra-Part I

Peter Minary, office: Room 367
email: peter.minary@cs.ox.ac.uk

- Chapter 1: Vectors and vector spaces

  (Week 1, Lectures 1-3)

- Chapter 2: Independence and orthogonality

  (Week 2, Lectures 4-6)

- Chapter 3: Matrices

  (Week 3, Lectures 7-9)

- Chapter4: Systems of linear equations

  (Week 4, Lectures 10-12)

Chapter 2

Independence and Orthogonality

2.1. Linear Independence and Basis

2.1.1. Linear combination of vectors

Definition 2.1.1 *If vector* $\mathbf{u}$ *in a vector space* V *can be expressed in the form*

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

where c_i, $i = 1, \ldots, n$ *are scalars, then* $\mathbf{u}$ *is called a linear combination of vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Example 2.1.2 *Suppose that* $\boldsymbol{V} = \mathbb{R}^2$,

$$\mathbf{u} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}, \mathbf{v_1} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}.$$

Then we have

$$\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} = \boldsymbol{a} \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} + \boldsymbol{b} \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}.$$

Therefore, $\mathbf{u} = \boldsymbol{a}\mathbf{v_1} + \boldsymbol{b}\mathbf{v_2}$.

Example 2.1.3 *Let* $\boldsymbol{V} = \mathbb{R}^3$ *and suppose we have vectors,*

$$\mathbf{u} = \begin{bmatrix} \mathbf{1} \\ \mathbf{-8} \\ \mathbf{12} \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} \mathbf{2} \\ \mathbf{-1} \\ \mathbf{3} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} \mathbf{5} \\ \mathbf{0} \\ \mathbf{4} \end{bmatrix}.$$

In order to express $\mathbf{u}$ *as a linear combination of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *we need* $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{R}$ *such that*

$$\boldsymbol{c}_1 \begin{bmatrix} \mathbf{2} \\ \mathbf{-1} \\ \mathbf{3} \end{bmatrix} + \boldsymbol{c}_2 \begin{bmatrix} \mathbf{5} \\ \mathbf{0} \\ \mathbf{4} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{-8} \\ \mathbf{12} \end{bmatrix}.$$

This translates to

$$\begin{aligned} \mathbf{2}\boldsymbol{c}_1 + \mathbf{5}\boldsymbol{c}_2 &= \mathbf{1} \\ -\boldsymbol{c}_1 &= \mathbf{-8} \\ \mathbf{3}\boldsymbol{c}_1 + \mathbf{4}\boldsymbol{c}_2 &= \mathbf{12} \end{aligned}$$

Substituting $\boldsymbol{c}_1 = \mathbf{8}$ *into the first equation gives* $\boldsymbol{c}_2 = \mathbf{-3}$*.*

Given that $c_1 = 8$ *and* $c_2 = -3$ *is consistent with the third equation we conclude that*

$$\begin{bmatrix} 1 \\ -8 \\ 12 \end{bmatrix} = 8 \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} + (-3) \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}.$$

Note, that it is not always possible to express a vector as a linear combination of other vectors.

Example 2.1.4 *We cannot express*

$$\mathbf{u} = \begin{bmatrix} \mathbf{4} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix}$$

as a linear combination of

$$\mathbf{v}_1 = \begin{bmatrix} \mathbf{2} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \textit{and} \quad \mathbf{v}_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{9} \end{bmatrix}.$$

Example 2.1.5 *Let $V = \mathbb{R}^3$ and suppose we have vectors,*

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}.$$

In order to write $\mathbf{u}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ we need to find $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ such that

$$\boldsymbol{c}_1 \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} + \boldsymbol{c}_2 \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

This translates to

$$\begin{aligned} 2\boldsymbol{c}_1 + 5\boldsymbol{c}_2 &= 1 \\ -\boldsymbol{c}_1 &= 1 \\ 3\boldsymbol{c}_1 + 4\boldsymbol{c}_2 &= -1 \end{aligned}$$

Substituting $\boldsymbol{c}_1 = -1$ into the first equation gives $\boldsymbol{c}_2 = \frac{3}{5}$, however these solutions are not consistent with the third equation so we conclude that such linear combination does not exist.

2.1.2. Application: linear combinations of Gaussian functions

Let $\mathbf{x}$ be any vector in $\mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^n$ a given vector. Then a Gaussian Radial Basis Function kernel or Gaussian kernel is defined as

$$\varphi(\mathbf{x}, \mathbf{c}) = \exp\left[-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2\sigma}\right].$$

This kernel represents a measure of similarity between vectors as 'closer' (defined by the squared norm of their distance) vectors have a larger Gaussian kernel value. Equivalently, the function describes the distance of any $\mathbf{x} \in \mathbb{R}^n$ from $\mathbf{c}$, which is often called the center.

Note that for $x, c \in \mathbb{R}$ we have the familiar function

$$\varphi(x, c) = \exp\left[-\frac{(x - c)^2}{2\sigma}\right].$$

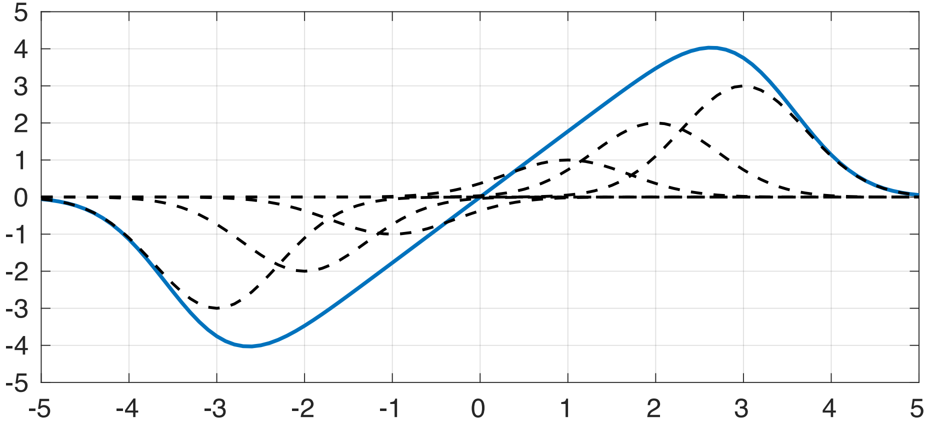
The Gaussian kernel is just one type of radial function, which satisfy

$$\varphi(\mathbf{x}, \mathbf{c}) = \varphi(\|\mathbf{x} - \mathbf{c}\|).$$

Most common application of Radial basis functions is their linear combination to express other functions, $\mathbf{y} = \sum_{i=1}^{m} \omega_i \varphi(\|\mathbf{x} - \mathbf{x}_i\|)$, where ω_i are scalars in this expansion.

Example 2.1.6 *Linear combination of one dimensional Gaussian kernels with centers,* $-3, -2, -1, 1, 2, 3$.

$$f(x) = \sum_{i=-3, i\neq 0}^{3} i \; \exp\left[(x-i)^2\right]$$



2.1.3. Spanning set and linear independence

Definition 2.1.7 *If* $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ *is a set of vectors in a vector space* V*, the set of all linear combinations of* $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ *is called the span of* $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ *and is denoted by* $span(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ *or* $span(S)$*. If* $V = span(S)$*, then* S *is called a spanning set for* V *and* V *is said to be spanned by* S*.*

Example 2.1.8 *Recall the previous example 2.1.2*

$$\begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

demonstrating that

$$\mathbb{R}^2 = span\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

Therefore,

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

is a spanning set for $\mathbb{R}^2$.

Example 2.1.9 *Is $\mathbb{R}^2$ also spanned by*

$$\boldsymbol{S} = \left\{ \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \right\}?$$

No, it is not, since

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \notin \boldsymbol{span(S)}.$$

On the other hand,

$$\boldsymbol{span(S)} = \left\{ \boldsymbol{c} \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \right\},$$

where $\boldsymbol{c} \in \mathbb{R}$.

Therefore, $\boldsymbol{S}$ is a spanning set for a line in $\mathbb{R}^2$.

Example 2.1.10 *The spanning set for* $\mathbb{R}^3$ *is*

$$\boldsymbol{S} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

since any vector $\mathbf{u} \in \mathbb{R}^3$ *with* $\mathbf{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3]^{\boldsymbol{T}}$ *can be written as*

$$\begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \boldsymbol{u}_3 \end{bmatrix} = \boldsymbol{u}_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{u}_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \boldsymbol{u}_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Example 2.1.11 $S = \{1, x, x^2, x^3\}$ *spans* $\mathcal{P}^3$ *because any polynomial function of* $p(x) = a + bx + cx^2 + dx^3$ *can be expressed in the form*

$$p(x) = p_0[1] + p_1[x] + p_2[x^2] + p_3[x^3],$$

where $p_0 = a$, $p_1 = b$, $p_2 = c$ *and* $p_3 = d$.

Definition 2.1.12 *A set of vectors* $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ *are said to be linearly independent if*

$$\boldsymbol{\alpha}_1 \mathbf{v}_1 + \boldsymbol{\alpha}_2 \mathbf{v}_2 + \cdots + \boldsymbol{\alpha}_k \mathbf{v}_k = \mathbf{0},$$

then $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \cdots = \boldsymbol{\alpha}_k = \mathbf{0}$. *If a set of not all zero* $\boldsymbol{\alpha}_i$*'s exist then the set is said to be linearly dependent.*

Example 2.1.13 *Using previous example 2.1.2*

$$\begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

we get

$$1 \begin{bmatrix} a \\ b \end{bmatrix} - a \begin{bmatrix} 1 \\ 0 \end{bmatrix} - b \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0.$$

Therefore,

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix} \right\}$$

is linearly dependent.

Example 2.1.14 *In $\mathcal{P}^2$, the set* $\{1+x+x^2, 1-x+3x^2, 1+3x-x^2\}$ *is linearly dependent, since*

$$2(1+x+x^2) - 1(1-x+3x^2) - 1(1+3x-x^2) = 0.$$

Example 2.1.15 *In $\mathcal{P}^n$ the set $\{1, x, x^2, \ldots, x^n\}$ is linearly independent.*
This is true because

$$p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n = 0$$

implies that the scalars are

$$p_0 = p_1 = p_2 = \cdots p_n = 0.$$

Example 2.1.16 *In $\mathcal{P}^2$, the set $\{1+x, x+x^2, 1+x^2\}$ is linearly independent.*
Suppose there exists α_1, α_2 and α_3, not all zero, such that

$$\alpha_1(1+x) + \alpha_2(x+x^2) + \alpha_3(1+x^2) = 0$$

for all $x \in \mathbb{R}$.
We now equate coefficients of each polynomial degree. The constants give

$$\alpha_1 + \alpha_3 = 0,$$

the linears give

$$\alpha_1 + \alpha_2 = 0,$$

and the quadratics give

$$\alpha_2 + \alpha_3 = 0.$$

This system of equations has the unique solution, $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Exercise 2.1.17 *Given independent vectors* $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, *show that the vectors* $\mathbf{u}_1 = \mathbf{v}_2 - \mathbf{v}_3$, $\mathbf{u}_2 = \mathbf{v}_1 - \mathbf{v}_3$ *and* $\mathbf{u}_3 = \mathbf{v}_2 - \mathbf{v}_1$ *are linearly dependent.*

Theorem 2.1.18 *Let* $S = \{\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_k\}$ *be a set with at least two elements* ($\boldsymbol{k \geq 2}$)*. Then* $\boldsymbol{S}$ *is linearly dependent if and only if one can express at least one vector as a linear combination of other vectors in* $\boldsymbol{S}$*.*

proof 2.1.19 $\Leftarrow$ *Assume that there is a vector we can express as a linear combination of the other vectors. We can always rearrange vectors so that the vector becomes* $\mathbf{v}_1$*. Then*

$$\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_k\mathbf{v}_k.$$

Therefore, the equation

$$\mathbf{0} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k.$$

have a nontrivial solution where at least one of the coefficients is nonzero, $c_1 = -1$*.*

$\Rightarrow$ *Suppose that* S *is linearly dependent. Then there is at least one nonzero among the scalars* $c_1, c_2, \ldots, c_k$ *in the equation*

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Suppose $c_1 \neq 0$ *(otherwise we can always rearrange the terms accordingly). Then we have*

$$\mathbf{v}_1 = \left(-\frac{c_2}{c_1}\right)\mathbf{v}_2 + \left(-\frac{c_3}{c_1}\right)\mathbf{v}_3 + \cdots + \left(-\frac{c_k}{c_1}\right)\mathbf{v}_k. \quad \square$$

Corollary 2.1.20 *$S = \{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if and only if $\mathbf{v}_1 = c\mathbf{v}_2$, where c is a scalar.*

Example 2.1.21

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \\ 10 \end{bmatrix}$$

so

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$$

with a nontrivial solution, with $c_1 = 1$ and $c_2 = -\frac{1}{2}$.

Therefore, $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent.

Example 2.1.22 *Recall we used $C[a, b]$ to denote the vector space of continuous functions defined on $[a, b]$. Let f, g in $C[-\pi, \pi]$ such that $f(x) = \sin(x)$ and $g(x) = \cos(x)$.*

Note, that

$$\cos(x) = \cos(-x)$$

and

$$\sin(-x) = -\sin(x)$$

therefore they cannot be scalar multiples of each other.

So they must be linearly independent!

Further note that $f, g \in C[a, b]$ are linearly independent if

$$c_1 f(x) + c_2 g(x) = 0$$

for all $x \in [a, b]$ implies that $c_1 = c_2 = 0$.

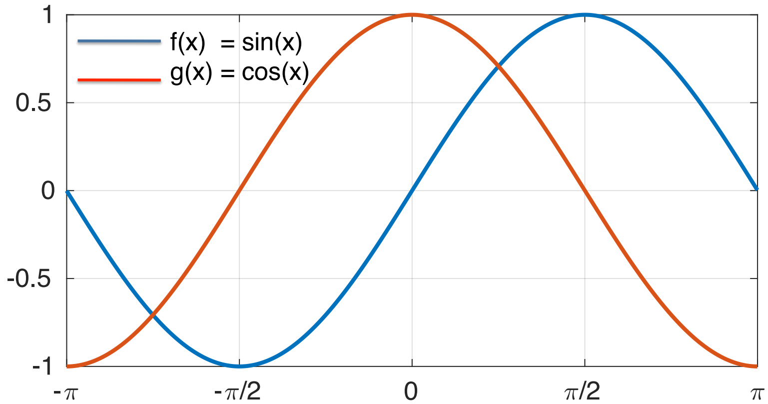
In our case suppose that

$$c_1 \sin(x) + c_2 \cos(x) = 0$$

for all $x \in [-\pi, \pi]$.

If $x = 0$ *we have* $\quad c_1 0 + c_2 1 = 0 \quad$ *leading to* $c_2 = 0$.
If $x = \frac{\pi}{2}$ *we have* $\quad c_1 1 + c_2 0 = 0 \quad$ *leading to* $c_1 = 0$.

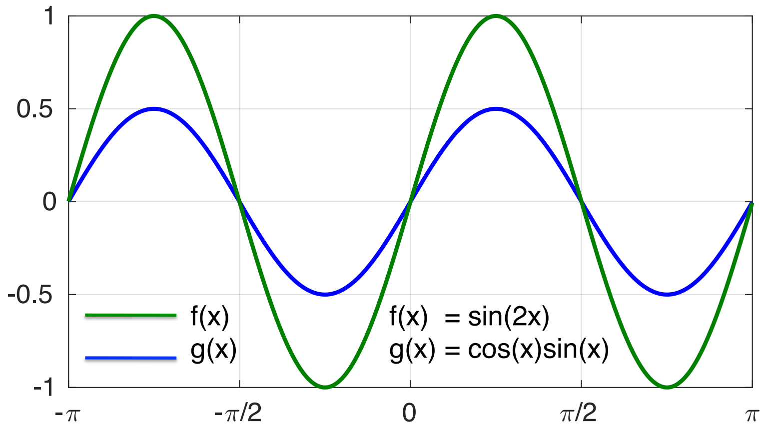
Therefore, $\cos(x)$ and $\sin(x)$ are linearly independent.



Example 2.1.23 *Let $f(x) = sin(2x)$ and $g(x) = sin(x)cos(x)$. Then*

$$c_1 f(x) + c_2 g(x) = 0$$

has the nontrivial solution $c_1 = 1$, $c_2 = -2$. The functions are linearly dependent.



Exercise 2.1.24 *Let* $V = span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$*, and suppose that one vector can be written as a linear combination of the other* $n-1$ *vectors. Prove that these* $n-1$ *vectors also form a spanning set for* V*.*

2.1.4. Basis for a vector space

Definition 2.1.25 *A basis for a vector space V is a set of linearly independent vectors that spans V.*
Suppose $S = \{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_n\}$ is a finite subset of a vector space V. Then S is a basis for V if

1. *The v_i are linearly independent.* *Not too many vectors*
2. $\boldsymbol{span(S) = V}$. *Not too few vectors*

In this case V is said to be finite dimensional.

Example 2.1.26 *Recall we showed that*

$$\boldsymbol{S} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a spanning set for $\mathbb{R}^3$. *Furthermore,*

$$\boldsymbol{c}_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{c}_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \boldsymbol{c}_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{0}$$

has only the trivial solution, $\boldsymbol{c}_1 = \boldsymbol{c}_2 = \boldsymbol{c}_3 = \mathbf{0}$, *so* $\boldsymbol{S}$ *is linearly independent.*

Given that $\boldsymbol{span}(\boldsymbol{S}) = \mathbb{R}^3$ *and* $\boldsymbol{S}$ *is linearly independent,*
$\boldsymbol{S}$ *is a basis for* $\mathbb{R}^3$.

Example 2.1.27 $S = \{1, x, x^2, x^3, \ldots, x^n\}$ *is a basis for* $\mathcal{P}^n$.
We already showed (in Example 2.1.15) that $S = \{1, x, x^2, x^3, \ldots, x^n\}$ *is linearly independent and we saw (in Example 2.1.11) that* $\{1, x, x^2, x^3\}$ *is a spanning set for* $\mathcal{P}^3$. *The latter conclusion can be straightforwardly generalized for* $\mathcal{P}^n$ *as any* $p(x) \in \mathcal{P}^n$ *can be written as a linear combination of vectors in* S.
Therefore, S *is a basis for* $\mathcal{P}^n$.

Theorem 2.1.28 *If* $S = \{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_n\}$ *is a basis for vector space* V *then any vector in* V *can be written as a unique linear combination of vectors in* S.

proof 2.1.29 *Since S spans V, any arbitrary vector $\mathbf{w}$ in V can be written as*

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \cdots + c_n\mathbf{v}_n.$$

Suppose that this representation is not unique and $\mathbf{w}$ can also be written as

$$\mathbf{w} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 \cdots + d_n\mathbf{v}_n.$$

Subtracting this second equation from the first yields

$$\mathbf{0} = (c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \ldots + (c_n - d_n)\mathbf{v}_n.$$

Since S is a basis it is linearly independent. Therefore, the above equation has only the trivial solution leading to: $c_i - d_i = 0$ for $i = 1, \ldots n$. Therefore, $\mathbf{w}$ has only one representation in basis S.

If $\boldsymbol{S} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis for vector space $\boldsymbol{V}$ then it is very easy to see that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n, \mathbf{u}\}$, where $\mathbf{u}$ is in $\boldsymbol{V}$, is linearly dependent!
A more general statement is also true.

Theorem 2.1.30 *Given that* $S = \{\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_n\}$ *is a basis for vector space* $\boldsymbol{V}$ *then any set that contain more than* $\boldsymbol{n}$ *vectors is linearly dependent.*

For Interest

proof 2.1.31 *Suppose that we compose $S' = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ from $m > n$ vectors in V. Then S' is linearly dependent if*

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \ldots a_m\mathbf{u}_m = \mathbf{0}$$

has a nontrivial solution, in which not all $a_i, i = 1, \ldots, m$ are zero. Given that S is a basis for V we can express each $\mathbf{u}_i$, $i = 1, \ldots m$ as a unique linear combination of vectors in S. Thus

$$\mathbf{u}_i = c_{1i}\mathbf{v}_1 + c_{2i}\mathbf{v}_2 + \cdots + c_{ni}\mathbf{v}_n \quad i = 1, \ldots m.$$

Substituting this set of equations into the previous equation yields

$$\begin{aligned} a_1(c_{11}\mathbf{v}_1 + c_{21}\mathbf{v}_2 + \cdots + c_{n1}\mathbf{v}_n) &\ + \\ a_2(c_{12}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{n2}\mathbf{v}_n) &\ + \ \cdots \\ + \ a_m(c_{1m}\mathbf{v}_1 + c_{2m}\mathbf{v}_2 + \cdots + c_{nm}\mathbf{v}_n) &\ = \ \mathbf{0}. \end{aligned}$$

By rearranging, we get

$$\begin{aligned} &(\boldsymbol{a_1c_{11}} + \boldsymbol{a_2c_{12}} + \cdots + \boldsymbol{a_mc_{1m}})\mathbf{v}_1 \ + \\ &(\boldsymbol{a_1c_{21}} + \boldsymbol{a_2c_{22}} + \cdots + \boldsymbol{a_mc_{2m}})\mathbf{v}_2 \ + \ \cdots \\ + \ &(\boldsymbol{a_1c_{n1}} + \boldsymbol{a_2c_{n2}} + \cdots + \boldsymbol{a_mc_{nm}})\mathbf{v}_n \ = \ \mathbf{0}. \end{aligned}$$

By choosing

$$\boldsymbol{b_j = a_1c_{j1} + a_2c_{j2} + \cdots + a_mc_{jm}, \quad j = 1, \ldots, n}$$

we get

$$\boldsymbol{b_1}\mathbf{v}_1 + \boldsymbol{b_2}\mathbf{v}_2 + \cdots + \boldsymbol{b_n}\mathbf{v}_n = \mathbf{0}.$$

Given that $\boldsymbol{S} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ *is linearly independent,*

$$\boldsymbol{b_1 = b_2 = \cdots b_n} = \mathbf{0}.$$

So we have

$$\boldsymbol{b_j = a_1 c_{j1} + a_2 c_{j2} + \cdots + a_m c_{jm} = 0, \ \ j = 1, \ldots, n}.$$

There are only $\boldsymbol{n}$ *equations but* $\boldsymbol{m > n}$ *variables* $\boldsymbol{a_i}$, $\boldsymbol{i = 1, \ldots m}$ *implying an infinite number of solutions. Therefore, there must be a solution apart from the trivial solution (*$\boldsymbol{a_i = 0}$*, i = 1, . . . , m). Therefore, for the equation*

$$\boldsymbol{a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \ldots a_m \mathbf{u}_m = 0}$$

there exists a solution for which at least one of the coefficients $\boldsymbol{a_1, \ldots, a_m}$ *has a nonzero value. Therefore,* $\boldsymbol{S'}$ *is linearly dependent.* $\square$

Theorem 2.1.32 *Any two bases for a vector space* $\boldsymbol{V}$ *contain the same number of vectors.*

proof 2.1.33 *Suppose both* $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ *and* $S' = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ *are both bases for vector space* V. *Theorem 2.1.30 implies that if* S *is a basis then* S' *cannot have more vectors otherwise it is not linearly independent (so it cannot be a basis). Therefore,* $m \leq n$. *By switching* S *and* S' *we can use the same argument to deduce that* $n \leq m$. *Therefore* $n = m$.

Definition 2.1.34 *If a vector space $\boldsymbol{V}$ has a basis with $\boldsymbol{n}$ vectors then $\boldsymbol{n}$ is called the dimension of $\boldsymbol{V}$ or $\boldsymbol{dim(V) = n}$.*

Example 2.1.35 *For some familiar vector spaces we have the following dimensions,*

$\boldsymbol{dim(\mathcal{P}^n) = n + 1}$,

$\boldsymbol{dim(\mathbb{R}^n) = n}$.

Theorem 2.1.36 *Let V be a vector space of dimension n. The following two statements are true:*

1. *If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set of vectors in V, then S is a basis for V.*
2. *If $V = span(S)$ then S is a basis for V.*

proof 2.1.37 *1. Let S be linearly independent and suppose that it does not form a basis. Then, there is a vector $\mathbf{u}$ in V such that $\mathbf{u} \notin span(S)$. Consequently the set $S' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{u}\}$ is linearly independent. By definition we know that $dim(V) = n$ implies that a basis for V has n vectors and due to Theorem 2.1.30, S' must be linearly dependent! This is a contradiction, therefore, S forms a basis for V.*

2. For the second part we let $span(S) = V$ and again assume that S does not form a basis. Then S is not linearly independent so $\mathbf{v}_i \in S$ can be expressed as a linear combination of other vectors. Without loss of generality we assume that $i = n$. Therefore, $span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}) = span(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = V$. But $n - 1$ vectors can span at most an $n - 1$ dimensional vector space. Therefore, S is linearly independent and forms a basis for V.

Exercise 2.1.38 *Suppose that* $\boldsymbol{W}$ *is a set of* $\boldsymbol{n} \in \mathbb{N}$ *nonzero vectors in a finite dimensional vector space* $\boldsymbol{V}$ *and that* $\boldsymbol{span}(\boldsymbol{W}) = \boldsymbol{V}$*. Prove that there exists* $\boldsymbol{W_b} \subseteq \boldsymbol{W}$ *such that* $\boldsymbol{W_b}$ *is a basis for* $\boldsymbol{V}$*.*

Exercise 2.1.39 *Suppose that* $\boldsymbol{W} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ *is a linearly independent set of vectors in a finite dimensional vector space* $\boldsymbol{V}$*. Prove that there exists a basis,* $\boldsymbol{W_b}$ *for* $\boldsymbol{V}$ *such that* $\boldsymbol{W} \subseteq \boldsymbol{W_b}$*.*

2.2. Orthogonal Vectors and Orthogonal Subspaces

Recall two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal if

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \mathbf{v} \cdot \mathbf{u} = \mathbf{v}^T \mathbf{u} = \mathbf{0}.$$

2.2.1. Orthogonality of vectors in vector spaces (optional)

The dot product in $\mathbb{R}^n$ enabled us to define length of vectors, unique angle between any two vectors in $\mathbb{R}^n$ and the latter enabled us to define orthogonality between vectors. A natural question arise whether we can extend these concepts to more common real vector spaces so that we could establish analog relationships between their vectors.

This is accomplished by the definition of the inner product of two vectors. Note, that the inner product is a generalization of the familiar dot product to vector spaces. In fact, dot product is just one of various inner products one can define on $\mathbb{R}^n$

Definition 2.2.1 *Suppose,* $\mathbf{u}, \mathbf{v},$ *and* $\mathbf{w}$ *are vectors in a vector space* $\boldsymbol{V}$*. Given any scalar* $\boldsymbol{c}$ *the inner product associates a real number, denoted by* $\langle \mathbf{u}, \mathbf{v} \rangle$ *with all pairs of vectors,* $\mathbf{u}$ *and* $\mathbf{v}$*. Furthermore,* $\langle \mathbf{u}, \mathbf{v} \rangle$ *satisfies*

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
3. $\boldsymbol{c}\langle \mathbf{u}, \mathbf{v} \rangle = \langle \boldsymbol{c}\mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq \mathbf{0}$, *and* $\langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{0}$ *if and only if* $\mathbf{v} = \mathbf{0}$.

A vector spaces with an inner product is often called inner product space.

Example 2.2.2 *Suppose* $\mathbf{u}$ *and* $\mathbf{v}$ *are two vectors in* $\mathbb{R}^n$ *and that the definition of the inner product is given by* $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$ *or equivalently* $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$ *if both* $\mathbf{u}$ *and* $\mathbf{v}$ *are column vectors,* $\mathbf{v}, \mathbf{u} \in \mathbb{R}^{n \times 1}$*. According to Theorem 1.1.6 all these properties are satisfied by the dot product. Therefore the dot product is an inner product.*

Exercise 2.2.3 *(optional) Suppose* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *and the function has the following definition*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} c_i u_i v_i,$$

where $\{c_i\}_{i=1}^n$ *are a set of positive constants. Does this function define an inner product?*

Exercise 2.2.4 *(optional) In the next example, suppose that* $\mathbf{u}$ *,* $\mathbf{v}$ *are in* $\mathbb{R}^3$ *and*

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 - u_3 v_3.$$

Does the above function define an inner product?

Example 2.2.5 *For the vector space*

$$C[a, b] = \{f \mid f : [a, b] \to \mathbb{R}, f \text{ continuous}\}.$$

of continuous functions defined on $[a, b]$ *we define*

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

This function satisfies all the axioms of the inner product.

For Interest

1.

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = \langle g, f\rangle$$

2.

$$\begin{aligned}\langle f, g + h\rangle &= \int_a^b f(x)(g(x) + h(x))dx \\ &= \int_a^b (f(x)g(x) + f(x)h(x))\, dx\end{aligned}$$

$$= \int_a^b f(x)g(x)dx + \int_a^b f(x)h(x)dx = \langle f, g\rangle + \langle f, h\rangle$$

3.

$$c\langle f, g\rangle = \int_a^b cf(x)g(x)dx = \int_a^b (cf(x))g(x)dx = \langle cf, g\rangle$$

4.

$$\langle g, g\rangle = \int_a^b g(x)g(x)dx = \int_a^b g^2(x)dx \geq 0$$

The powerful concept of the inner product enables us to characterize vectors in vector spaces with an inner product. Let $\boldsymbol{V}$ be such a vector space. Analogously to vectors in $\mathbb{R}^{n}$ the length or norm of a vector $\mathbf{u} \in \boldsymbol{V}$ is given by

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Exercise 2.2.6 *(optional) Let $f, g \in C[-1, 1]$ and the inner product defined as*

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) dx.$$

Suppose $f(x) = x$. Calculate $\|f\|$.

Once the concept of the length is established all important inequalities discussed for vectors in $\mathbb{R}^n$ follows.

Theorem 2.2.7 *If $\mathbf{u}$ and $\mathbf{v}$ are vectors in V then the following inequalities hold:*

1. $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$ *Cauchy-Schwartz Inequality*
2. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ *Triangle Inequality*

The proof of these inequalities follows line by line to the proof of the corresponding inequalities for the case of $\mathbb{R}^n$ by simply replacing $\mathbf{u} \cdot \mathbf{v}$ with $\langle \mathbf{u}, \mathbf{v} \rangle$.

Exercise 2.2.8 *(optional) Let $f, g \in C[0, 1]$, $f(x) = x, g(x) = \exp(x)$ and the inner product defined as*

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx.$$

Show that the Cauchy-Schwartz inequality holds for vectors f and g.

The Cauchy-Schwartz Inequality implies that

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \leq 1,$$

which enables the definition of a unique angle between vectors $\mathbf{u}$ and $\mathbf{v}$ in $\boldsymbol{V}$.

Furthermore, we say that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{0}.$$

In this last expression $\mathbf{u}$ and $\mathbf{v}$ may refer to vectors in vector spaces including both $\mathbb{R}^n$ and $C[a, b]$.

In the remainder of the document however we will use $\mathbf{u} \cdot \mathbf{v}$ (or equivalently $\mathbf{u}^T\mathbf{v}$) to emphasize that we refer to the dot (inner) product of vectors in $\mathbb{R}^n$.

2.2.2. Orthogonal subspaces

Definition 2.2.9 *Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ are said to be orthogonal subspaces if every vector in $\boldsymbol{V}$ is orthogonal to every vector in $\boldsymbol{W}$.*

Example 2.2.10 *In $\mathbb{R}^{\mathbf{3}}$ the subspace spanned by one vector (a line) can be orthogonal to another line, or a plane. A subspace spanned by two vectors (a plane) cannot be orthogonal to another plane!*

Definition 2.2.11 *Given a subspace $\boldsymbol{V} \subset \mathbb{R}^{\boldsymbol{n}}$, the space of all vectors orthogonal to $\boldsymbol{V}$ is called the orthogonal complement of $\boldsymbol{V}$. It is denoted $\boldsymbol{V}^{\perp}$, we say '$\boldsymbol{V}$ perp'.*

If for subspaces $\boldsymbol{V}, \boldsymbol{W} \subset \mathbb{R}^{\boldsymbol{n}}$ we have $\boldsymbol{W} = \boldsymbol{V}^{\perp}$ then $\boldsymbol{V} = \boldsymbol{W}^{\perp}$ and $\boldsymbol{dim}\ \boldsymbol{V} + \boldsymbol{dim}\ \boldsymbol{W} = \boldsymbol{n}$. In other words we have decomposed the whole space into two perpendicular parts. For every $\mathbf{x} \in \mathbb{R}^{\boldsymbol{n}}$ we have the vectors $\mathbf{v} \in \boldsymbol{V}$ (projection onto $\boldsymbol{V}$) and $\mathbf{w} \in \boldsymbol{W}$ (projection onto $\boldsymbol{W}$) such that $\mathbf{x} = \mathbf{v} + \mathbf{w}$.

Next sections we will introduce definitions and statements applicable for vectors in $\mathbb{R}^{\boldsymbol{n}}$ and we will use Remarks to mention the extension of these concepts to vector spaces in general.

These Remarks will serve as optional material.

2.2.3. Orthogonal and orthonormal sets

Definition 2.2.12 *The set of vectors form an orthogonal set* $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ *if* $\mathbf{q}_i \cdot \mathbf{q}_j = \mathbf{q}_i^T \mathbf{q}_j = 0$ *for all* $i \neq j$.

Theorem 2.2.13 *The vectors of an orthogonal set are linearly independent.*

proof 2.2.14 *Let* $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ *be nonzero vectors and assume that the set is linearly dependent. Then there exist* $\mathbf{q}_j \in \{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ *such that*

$$\mathbf{q}_j = \sum_{i=1, i\neq j}^{k} c_i \mathbf{q}_i$$

Given that $\mathbf{q}_j$ *is nonzero,* $\mathbf{q}_j \cdot \mathbf{q}_j \neq 0$*. However, the sum*

$$\mathbf{q}_j \cdot \mathbf{q}_j = \mathbf{q}_j^T \mathbf{q}_j = \sum_{i=1, i\neq j}^{k} c_i \left(\mathbf{q}_j^T \mathbf{q}_i\right) = 0$$

is zero because the vectors in $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ *are orthogonal. This contradicts the initial assumption that vectors are nonzero. Thus, vectors are linearly independent.*

Definition 2.2.15 *A set of vectors* $\{\mathrm{q}_1, \mathrm{q}_2, \cdots, \mathrm{q}_k\}$ *are orthonormal if they are orthogonal and* $\mathrm{q}_i \cdot \mathrm{q}_i = 1$ *for all* $i = 1, \ldots, k$*. In this case* $\{\mathrm{q}_1, \mathrm{q}_2, \cdots, \mathrm{q}_k\}$ *is called an orthonormal set.*

Remark 2.2.16 *(optional) In general the set of vectors* $\{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_k\}$ *in a vector space forms an orthogonal set if* $\langle \mathrm{s}_i, \mathrm{s}_j \rangle = 0$ *for all* $i \neq j$*. If* $\langle \mathrm{s}_i, \mathrm{s}_i \rangle = 1$ *also holds for all* $i = 1, \ldots, k$ *then* $\{\mathrm{s}_1, \mathrm{s}_2, \cdots, \mathrm{s}_k\}$ *is an orthonormal set.*

2.2.4. Decomposition into orthogonal components

Assume that $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ is an orthogonal set of vectors and $\mathbf{u}$ is an arbitrary vector. Then

$$\mathbf{v} = \mathbf{u} - \frac{\mathbf{q}_1^T \mathbf{u}}{\mathbf{q}_1^T \mathbf{q}_1}\mathbf{q}_1 - \cdots - \frac{\mathbf{q}_k^T \mathbf{u}}{\mathbf{q}_k^T \mathbf{q}_k}\mathbf{q}_k$$

is orthogonal to $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$.

This is true since for any $\mathbf{q}_i, i = 1, ..., k$, we have

$$\mathbf{q}_i^T \mathbf{v} = \mathbf{q}_i^T \mathbf{u} - \frac{\mathbf{q}_1^T \mathbf{u}}{\mathbf{q}_1^T \mathbf{q}_1} \mathbf{q}_i^T \mathbf{q}_1 - \cdots - \frac{\mathbf{q}_k^T \mathbf{u}}{\mathbf{q}_k^T \mathbf{q}_k} \mathbf{q}_i^T \mathbf{q}_k$$

$$\mathbf{q}_i^T \mathbf{v} = \mathbf{q}_i^T \mathbf{u} - \frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} (\mathbf{q}_i^T \mathbf{q}_i) = \mathbf{q}_i^T \mathbf{u} - \mathbf{q}_i^T \mathbf{u} = 0$$

Therefore, we can decompose $\mathbf{u}$ into $\boldsymbol{k+1}$ orthogonal components:

$$\mathbf{u} = \mathbf{v} + \sum_{i=1}^{k} \frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

where $\mathbf{v}$ is the component of $\mathbf{u}$ that is orthogonal to $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ and

$$\frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

is the component of $\mathbf{u}$ that is along the direction of $\mathbf{q}_i$.

We know that $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ linearly independent vectors form a basis in $\mathbb{R}^k$. Therefore, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$ and then $\mathbf{v} = \mathbf{0}$ must hold and $\mathbf{u}$ is decomposed into orthogonal components along directions $\mathbf{q}_i$.

$$\mathbf{u} = \sum_{i=1}^{k} \frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

We may also write this sum in the form

$$\mathbf{u} = \sum_{i=1}^{k} \mathbf{proj}_{\mathbf{q}_i} \mathbf{u},$$

if we use a notation

$$\mathbf{proj}_{\mathbf{q}_i} \mathbf{u} = \frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

for the projection of $\mathbf{u}$ along $\mathbf{q}_i$.

If $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ is an orthonormal set of vectors and $\mathbf{u}$ is a vector in $\mathbb{R}^k$ then the sum

$$\mathbf{u} = \sum_{i=1}^{k} \frac{\mathbf{q}_i^T \mathbf{u}}{\mathbf{q}_i^T \mathbf{q}_i} \mathbf{q}_i$$

reduces to

$$\mathbf{u} = \sum_{i=1}^{k} (\mathbf{q}_i^T \mathbf{u}) \mathbf{q}_i,$$

since $\mathbf{q}_i^T \mathbf{q}_i = 1$ for all $i = 1, \ldots, k$. Note that

$$(\mathbf{q}_i^T \mathbf{u}) \mathbf{q}_i$$

is the component of $\mathbf{u}$ that is along the direction of $\mathbf{q}_i$.

Remark 2.2.17 *(optional) If set of vectors $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_k\}$ in a vector space forms an orthogonal set and $\mathbf{u}$ is a vector in the vector space then*

$$\frac{\langle \mathbf{s}_i, \mathbf{u} \rangle}{\langle \mathbf{s}_i, \mathbf{s}_i \rangle} \mathbf{s}_i$$

is the component of $\mathbf{u}$ along $\mathbf{s}_i$.

If $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_k\}$ is an orthonormal set, then the component of $\mathbf{u}$ along $\mathbf{s}_i$ is

$$\langle \mathbf{s}_i, \mathbf{u} \rangle \mathbf{s}_i.$$

Exercise 2.2.18 *Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$. Prove that*

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} |\mathbf{v} \cdot \mathbf{u}_i|^2.$$

Exercise 2.2.19 *Let $\mathbf{s}$ and $\mathbf{u}$ be vectors in $\mathbb{R}^2$ with an angle φ between them. Use geometric considerations to derive a formula for the projection of $\mathbf{u}$ along $\mathbf{s}$.*

2.3. Gram-Schmidt orthogonalization

Let $\boldsymbol{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ be a set of linearly independent vectors in $\mathbb{R}^m$ and $\boldsymbol{m} \geq \boldsymbol{n}$. Then the goal would be to obtain an orthonormal set of vectors, $\boldsymbol{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$, such that $\boldsymbol{span}(\boldsymbol{Q}) = \boldsymbol{span}(\boldsymbol{X})$.

2.3.1. The basic idea

First we look at a simple example, when $\boldsymbol{n} = \mathbf{3}$. Starting from $\boldsymbol{X}$ we can generate the orthonormal set, $\boldsymbol{Q}$ by executing the following steps:

1. First we normalize $\mathbf{x}_1$.

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1.$$

2. Now we need to extract the $\mathbf{q}_1$ component of $\mathbf{x}_2$ to produce $\mathbf{q}_2$ and then normalize the result

$$\mathbf{q}_2 = \frac{\mathbf{x}_2 - (\mathbf{q}_1^T\mathbf{x}_2)\mathbf{q}_1}{\|\mathbf{x}_2 - (\mathbf{q}_1^T\mathbf{x}_2)\mathbf{q}_1\|}.$$

3. Given that the original set was linearly independent $\mathbf{x_3}$ cannot lie in the plane spanned by $\mathbf{x_1}$ and $\mathbf{x_2}$. We now define

$$\mathbf{q_3} = \frac{\mathbf{x_3} - (\mathbf{q_1^{\mathit{T}} x_3})\mathbf{q_1} - (\mathbf{q_2^{\mathit{T}} x_3})\mathbf{q_2}}{\|\mathbf{x_3} - (\mathbf{q_1^{\mathit{T}} x_3})\mathbf{q_1} - (\mathbf{q_2^{\mathit{T}} x_3})\mathbf{q_2}\|}.$$

If we have more vectors we carry on in the same manner. This leads to the Gram-Schmidt process:

Given linearly independent vectors $\mathbf{x_1}, \mathbf{x_2}, \ldots \mathbf{x_n}$, the j^{th} orthonormal vector is given by

$$\mathbf{q}_j = \frac{\mathbf{x}_j - \sum_{i=1}^{j-1}(\mathbf{q}_i^T \mathbf{x}_j)\mathbf{q}_i}{\left\| \mathbf{x}_j - \sum_{i=1}^{j-1}(\mathbf{q}_i^T \mathbf{x}_j)\mathbf{q}_i \right\|}.$$

Note, that after successively using this formula up to the calculation of the j^{th} orthonomal vector we have

$$span(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_j) = span(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_j) \quad j = 1, .., n.$$

Remark 2.3.1 *(optional) Note that the Gram-Schmidt process can be extended to vector spaces with an inner product:*

Given linearly independent vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n$*, the* j^{th} *orthonormal vector is given by*

$$\mathbf{q}_j = \frac{\mathbf{s}_j - \sum_{i=1}^{j-1} \langle \mathbf{q}_i, \mathbf{s}_j \rangle \mathbf{q}_i}{\left\| \mathbf{s}_j - \sum_{i=1}^{j-1} \langle \mathbf{q}_i, \mathbf{s}_j \rangle \mathbf{q}_i \right\|}.$$

2.3.2. The classical Gram-Schmidt algorithm

Starting from an initial set $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ of linearly independent vectors, the following pseudocode illustrates the steps in the classical Gram-Schmidt algorithm used to generate the orthonormal set, $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$.

Classical Gram-Schmidt

for $j = 1$ **to** n
    $\mathbf{v}_j = \mathbf{x}_j$
    for $i = 1$ **to** $j - 1$
        $s_{ij} = \mathbf{q}_i^T \mathbf{x}_j$
        $\mathbf{v}_j = \mathbf{v}_j - s_{ij}\mathbf{q}_i$
    $\mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\|$

The j^{th} step (main loop) of the algorithm implements the following definition of $\mathbf{v}_j$

$$\mathbf{v}_j = \mathbf{x}_j - (\mathbf{q}_1^T \mathbf{x}_j)\mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{x}_j)\mathbf{q}_2 - \cdots - (\mathbf{q}_{j-1}^T \mathbf{x}_j)\mathbf{q}_{j-1}$$

and we get $\mathbf{q}_j$ by normalizing $\mathbf{v}_j$.

Upon completion of the j^{th} step $\mathbf{q}_j$

- is normalized,
- is orthogonal to $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_{j-1}\}$, and
- $\mathbf{q}_j \in \boldsymbol{span}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_j)$.

This classical algorithm provides an excellent example demonstrating the steps of an orthogonalization methods.

2.3.3. The modified Gram-Schmidt algorithm (optional)

Compared to the classical Gram-Schmidt algorithm, the modified Gram-Schmidt algorithm is less sensitive to rounding errors on a computer. Fortunately, a very minor modification such as replacing line 4 with line 4' in

Classical Gram-Schmidt

1: **for** $j = 1$ **to** n

2: $\quad \mathbf{v}_j = \mathbf{x}_j$

3: $\quad$ **for** $i = 1$ **to** $j - 1$

4: $\qquad s_{ij} = \mathbf{q}_i^T \mathbf{x}_j$

[4': $\qquad s_{ij} = \mathbf{q}_i^T \mathbf{v}_j$]

5: $\qquad \mathbf{v}_j = \mathbf{v}_j - s_{ij}\mathbf{q}_i$

6: $\quad \mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\|$

we obtain the modified Gram-Schmidt algorithm:

Modified Gram-Schmidt

for $j = 1$ **to** n

$\quad \mathbf{v}_j = \mathbf{x}_j$

$\quad$ **for** $i = 1$ **to** $j - 1$

$\quad\quad s_{ij} = \mathbf{q}_i^T \mathbf{v}_j$

$\quad\quad \mathbf{v}_j = \mathbf{v}_j - s_{ij}\mathbf{q}_i$

$\quad \mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\|$

The $\boldsymbol{j}^{th}$ step (main loop) of the algorithm implements the following definition of $\mathbf{v}_j$

$$\begin{aligned}
\mathbf{v}_j^{(1)} &= \mathbf{x}_j \\
\mathbf{v}_j^{(2)} &= \mathbf{v}_j^{(1)} - \left(\mathbf{q}_1^T \mathbf{v}_j^{(1)}\right) \mathbf{q}_1 \\
&\cdot \\
&\cdot \\
\mathbf{v}_j^{(j)} &= \mathbf{v}_j^{(j-1)} - \left(\mathbf{q}_{j-1}^T \mathbf{v}_j^{(j-1)}\right) \mathbf{q}_{j-1}.
\end{aligned}$$

At the end, we get $\mathbf{q}_j$ by normalizing $\mathbf{v}_j$.

To summarize: the main differences in the classical Gram-Schmidt (CGS) and modified Gram-Schmidt (MGS) algorithms are

Procedure Gram-Schmidt

```
for j = 1 to n
    v_j = x_j
    for i = 1 to j - 1
        s_ij = q_i^T x_j        CGS
        s_ij = q_i^T v_j        MGS
        v_j = v_j - s_ij q_i
    q_j = v_j / ||v_j||
```

$$
\begin{array}{ll}
\textbf{for } j = 1 \textbf{ to } n & \\
\quad \mathbf{v}_j = \mathbf{x}_j & \\
\quad \textbf{for } i = 1 \textbf{ to } j-1 & \\
\qquad s_{ij} = \mathbf{q}_i^T \mathbf{x}_j & \textit{CGS} \\
\qquad s_{ij} = \mathbf{q}_i^T \mathbf{v}_j & \textit{MGS} \\
\qquad \mathbf{v}_j = \mathbf{v}_j - s_{ij}\mathbf{q}_i & \\
\quad \mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\| &
\end{array}
$$

At the j^{th} step in the main loop we have

(Classical Gram-Schmidt)

$$\mathbf{v}_j = \mathbf{x}_j - (\mathbf{q}_1^T \mathbf{x}_j)\mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{x}_j)\mathbf{q}_2 - \cdots - (\mathbf{q}_{j-1}^T \mathbf{x}_j)\mathbf{q}_{j-1}$$

(Modified Gram-Schmidt)

$$\begin{aligned}
\mathbf{v}_j^{(1)} &= \mathbf{x}_j \\
\mathbf{v}_j^{(2)} &= \mathbf{v}_j^{(1)} - \left(\mathbf{q}_1^T \mathbf{v}_j^{(1)}\right) \mathbf{q}_1 \\
&\cdot \\
&\cdot \\
\mathbf{v}_j^{(j)} &= \mathbf{v}_j^{(j-1)} - \left(\mathbf{q}_{j-1}^T \mathbf{v}_j^{(j-1)}\right) \mathbf{q}_{j-1}.
\end{aligned}$$

and we get $\mathbf{q}_j$ by normalizing $\mathbf{v}_j$.

The modified Gram-Schmidt algorithm successively applies the orthogonalization with respect to single vectors already available in the orthonormal set. Therefore, once $\mathbf{q}_i$ is known, it could be applied to all $\mathbf{v}_j^{(i)}$ for $j > i$. With these considerations, the efficient implementation of the modified Gram-Schmidt algorithm is

Modified Gram-Schmidt

for $i = 1$ **to** n
    $\mathbf{v}_i = \mathbf{x}_i$
for $i = 1$ **to** n
    $\mathbf{q}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$
    for $j = i + 1$ **to** n
        $s_{ij} = \mathbf{q}_i^T \mathbf{v}_j$
        $\mathbf{v}_j = \mathbf{v}_j - s_{ij} \mathbf{q}_i$

Operation Count

We obtain the cost of the algorithm, by counting the number of "flops" (floating points operations) such as '$+$', '-', '.', '/', '$\sqrt{\ }$' and assuming that $\mathbf{x}_i, \mathbf{q}_i, i = 1, \ldots, n$ are vectors in $\mathbb{R}^m$.

Modified Gram-Schmidt

```
for i = 1 to n
    v_i = x_i
for i = 1 to n
    v_l = ||v_i||            ← m '.' and m-1 '+' and 1 '√'
    q_i = v_i / v_l          ← m '/'
    for j = i + 1 to n
        s_ij = q_i^T v_j     ← m '.' and m-1 '+'
        v_j = v_j - s_ij q_i ← m '.' and m '-'
```

The sum of additions: $'+'$

$$\begin{aligned}
S_+ &= \sum_{i=1}^{n}\left(m-1+\sum_{j=i+1}^{n}(m-1)\right) \\
&= \sum_{i=1}^{n}(m-1)+\sum_{i=1}^{n}\sum_{j=i+1}^{n}(m-1) \\
&= n(m-1)+(m-1)\sum_{i=1}^{n}(n-i)=(m-1)\left(n+\sum_{i=1}^{n-1}i\right) \\
&= (m-1)\left(n+\frac{n(n-1)}{2}\right)=\frac{1}{2}(m-1)(n^2+n) \\
S_- &= \sum_{i=1}^{n}\sum_{j=i+1}^{n}m=m\sum_{i=1}^{n-1}i=\frac{1}{2}m(n^2-n)
\end{aligned}$$

$$S_. = \sum_{i=1}^{n}\left(m+\sum_{j=i+1}^{n}2m\right)=nm+\sum_{i=1}^{n}\left(\sum_{j=i+1}^{n}2m\right)$$

$$
\begin{aligned}
&= nm + 2m\sum_{i=1}^{n}(n-i) = nm + 2m\sum_{i=1}^{n-1} i \\
&= nm + 2m\frac{n(n-1)}{2} = nm + m(n^2 - n) = mn^2 \\
S_/ &= \sum_{i=1}^{n} m = nm \\
S_{total} &= S_+ + S_- + S_. + S_/ \\
&= \frac{1}{2}(m-1)(n^2+n) + \frac{1}{2}m(n^2-n) + mn^2 + nm \\
&= 2mn^2 - \frac{1}{2}(n^2+n) + nm \sim 2mn^2.
\end{aligned}
$$

Here $\sim$ indicates the leading term in case $n, m \to \infty$.

2.3.4. Orthonormal matrix

A collection of orthogonal column vectors $\mathbf{q}_i \in \mathbb{R}^m, \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}$, can be used to compose an $\boldsymbol{m} \times \boldsymbol{n}$ matrix:

$$\left[\begin{array}{c|c|c} & & \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ & & \end{array}\right] \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{n}}.$$

A matrix with orthonormal columns will be denoted by $\boldsymbol{Q}$ and is called an orthonormal matrix.

Example 2.3.2 *The standard basis* $\mathrm{e}_1, \mathrm{e}_2, \ldots \mathrm{e}_n$ *for* $\mathbb{R}^n$*.*

$$\left[\begin{array}{c|c|c} & & \\ \mathbf{e}_1 & \cdots & \mathbf{e}_n \\ & & \end{array}\right] \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{n}}$$

2.3.5. Orthogonal polynomials (optional)

The vector space can be defined based on a set of polynomials with maximum degree n or less. For example, vector space

$$\mathcal{P}^n = span(\{1, x, x^2, \cdots, x^n\})$$

so that any $p(x) \in \mathcal{P}^n$ can be expressed as

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

Note, that the elements of the spanning set $\{1, x, \cdots, x^n\}$ are vectors. Analogously, to the dot product $(\mathbf{v} \cdot \mathbf{u})$, here we can define the inner product, between any two vectors, $f(x)$ and $g(x)$ in the set as

$$\langle f, g \rangle = \int_a^b f(x) g(x) dx,$$

where $[a, b]$ is a given interval and $f(x)$, $g(x)$ are orthogonal if $\langle f, g \rangle = 0$. Furthermore, f is normalized if $\|f\| = \langle f, f \rangle = 1$.

With respect to this inner product the Gram-Schmidt procedure enables us to generate orthogonal polynomials on $[a, b]$, which we set to $[-1, 1]$.

If $\psi_j(x) = x^j, j = 0, 1, 2, \ldots$ then we can obtain $\{\varphi_j(x)\}$, orthogonal set by applying the Gram-Schmidt procedure.

$$
\begin{aligned}
v_0(x) &= \psi_0(x) \\
\varphi_0(x) &= \frac{v_0(x)}{\|v_0\|} = \frac{1}{\langle 1, 1\rangle^{1/2}} = \frac{1}{\left[\int_{-1}^{1} 1 ds\right]^{1/2}} = \frac{1}{\sqrt{2}} \\
v_1(x) &= \psi_1(x) - \langle \varphi_0, \psi_1\rangle \varphi_0 = x - \left\langle \frac{1}{\sqrt{2}}, x \right\rangle \frac{1}{\sqrt{2}} \\
&= x - \frac{1}{2}\int_{-1}^{1} x \, ds = x - \frac{1}{2} 0 = x \\
\varphi_1(x) &= \frac{v_1(x)}{\|v_1\|} = \frac{x}{\langle x, x\rangle^{1/2}} = \frac{x}{\left[\int_{-1}^{1} s^2 ds\right]^{1/2}} = \sqrt{\frac{3}{2}} x.
\end{aligned}
$$

We may consider an analog problem using an approach that only involves vectors in $\mathbb{R}^m$, $m \in \mathbb{N}$.

Based on $\psi_j(x) = x^j, j = 0, 1, 2, ...n$, let us define the vectors $\mathrm{x}_j \in \mathbb{R}^{2p+1}$, $j = 0, ..., n$ and $p \in \mathbb{N}$ such that

$$\mathrm{x}_j = \begin{bmatrix} \psi_j(x_1) \\ \psi_j(x_2) \\ \psi_j(x_3) \\ \cdot \\ \cdot \\ \psi_j(x_{2p+1}) \end{bmatrix} = \begin{bmatrix} \psi_j(-1) \\ \psi_j(-1+\delta) \\ \psi_j(-1+2\delta) \\ \cdot \\ \cdot \\ \psi_j(1) \end{bmatrix} \qquad \delta = \frac{1}{p}$$

and p controls the number of points the monomials $x^0, x^1, \ldots x^n$ are evaluated at points $x_1, x_2, \ldots, x_{2p+1}$ discretizing the interval $[-1, 1]$. In our case, we choose $p = 100$ so that we have a set $\{\mathrm{x}_j\}_{j=0}^n$ of $n+1$ vectors in $\mathbb{R}^{201}$.

By applying the Gram-Schmidt method on the set of vectors $\{\mathrm{x}_j\}_{j=0}^n$, $\mathrm{x}_j \in \mathbb{R}^{2p+1}$, we could obtain the orthogonal set $\{\mathrm{q}_j\}_{j=0}^n$, $\mathrm{q}_j \in \mathbb{R}^{2p+1}$.

Note, that $\{\mathrm{q}_j\}_{j=0}^n$ essentially represent orthogonal polynomials evaluated at $2p+1$ equidistant grid points $x_1, x_2, \ldots, x_{2p+1}$ on $[-1, 1]$.

In particular, we are interested in a set of orthogonal polynomials, $P_0, P_1, \ldots P_n$ that obey $P_i(1) = 1, i = 0, \ldots n$. Using proper scaling of vectors in $\{\mathrm{q}_j\}_{j=0}^n$ we can produce the vectors

$$\mathrm{p}_j = \frac{1}{q_{j(2p+1)}} \mathrm{q}_j, \quad j = 1, \ldots, n,$$

that represent the desired polynomials at the $2p+1$ equidistant grid points on $[-1, 1]$. These polynomials are called the Legendre polynomials.

Legendre polynomials

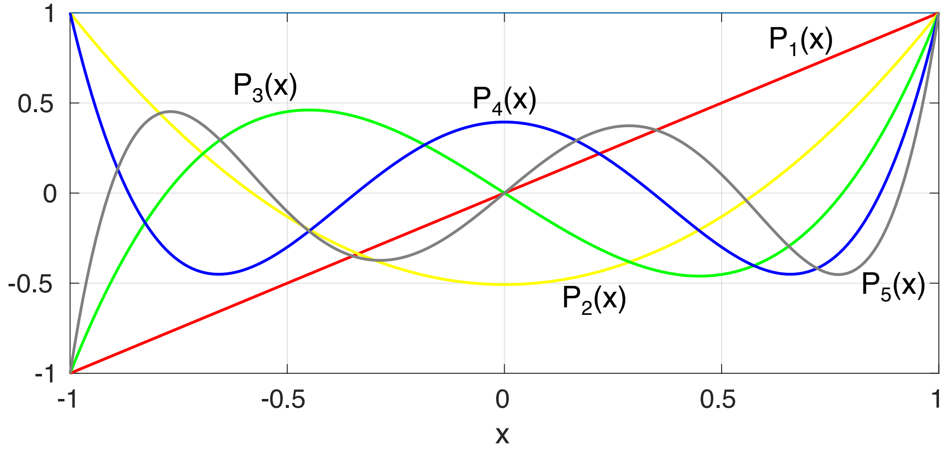
The first few Legendre polynomials are listed below.

$$
\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= \frac{1}{2}(3x^2 - 1) \\
P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\
P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3) \\
P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x) \\
&\vdots
\end{aligned}
$$

Furthermore, the Legendre polynomials are orthogonal as

$$\langle P_i, P_j \rangle = \int_{-1}^{1} P_i(x) P_j(x) dx = \frac{2}{2i+1} \delta_{ij},$$

where $\delta_{ij} = 1$ if $i = j$ otherwise $\delta_{ij} = 0$.



2.4. Numerical experiments (optional)

2.4.1. The implementation of the Gram-Schmidt algorithm (optional)

The figure shows the Gram-Schmidt algorithm as implemented in MATLAB.

```
function [Q] = clgs(X)
% [Q] = clgs(X)
% Gram-Schmidt orthogonalization
% X is an m x n matrix (m>=n)
% Q is an m x n matrix with orthogonal columns
n = size(X,2);
for j = 1:n
    v = X(:,j);
    for i = 1:j-1
        s(i,j) = Q(:,i)'*X(:,j);
        v = v - s(i,j)*Q(:,i);
    end
    Q(:,j) = v/norm(v);
end
```

The set of $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ linearly independent vectors in $\mathbb{R}^m$ $(m \geq n)$ form the columns of an m by n matrix X and the resulting m by n matrix, Q has orthogonal columns, with vectors $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$.
In the figure, $Q(:, i)$ denotes the i^{th} column of Q, in our notation, $\mathbf{q}_j$, $Q(:, i)' * X(:, j)$ is the dot product, in our notation, $\mathbf{q}_i^T \mathbf{x}_j$. Finally, $norm(v)$ is the length of the vector, $||v||$ in our notation.

The next panel shows the modified Gram-Schmidt algorithm as implemented in MATLAB.

```
function [Q] = mgs(X)
% modified Gram-Schmidt orthogonalization
% X is an m x n matrix (m>=n)
% Q is an m x n matrix with orthogonal columns
n = size(X,2);
   V = X;
   for i = 1:n
      Q(:,i) = V(:,i)/norm(V(:,i));
      for j = i+1:n
          s(i,j) = Q(:,i)'*V(:,j);
          V(:,j) = V(:,j) - s(i,j)*Q(:,i);
      end
   end
```

2.4.2. Application of the Gram-Schmidt algorithm (optional)

The Gram-Schmidt algorithms can be applied to compute Legendre polynomials on the interval $[-1, 1]$. The figure shows how to achieve this in MATLAB using the previous functions we defined.

```
function [Q] = legendpoly(n,p,l)
% Q = legendpoly(n,p,l)
% Compute Legendre polynomials on the [-1:1] interval
% n : up to nth order
% p : at 2p + 1 equidistant grid points
% l: using different orthogonalization methods
% l = 1 --> Classical Gram-Schmidt
% l = 2 --> Modified Gram-Schmidt
%
x = (-p:p)'/p;
for j=1:n+1
    X(:,j) = x.^(j-1);
end
 if l == 1
   [Q] = clgs(X);
 elseif l == 2
   [Q] = mgs(X);
 end
scale = Q(2*p+1,:);
Q = Q*diag(1./scale);

 plot(x,Q);
 title('Legendre Polynomials on [-1,1]');
 xlabel('x');
 ylabel(['P_i(x), i=0,...,' num2str(n) '']);
```