Math 524: Linear Algebra

Notes #6 — Inner Product Spaces

Peter Blomgren
⟨blomgren@sdsu.edu⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Fall 2021
(Revised: December 7, 2021)

Outline

Student Learning Targets, and Objectives 1 of 2

Target Inner Product Spaces, Cauchy–[Bunyakovsky]–Schwarz Inequality

- **Objective** Be able to state the Definitions and Properties of Inner Products, Norms, and Inner Product Spaces
- **Objective** Be able to use the Cauchy–[Bunyakovsky]–Schwarz inequality to show a variety of inequalities

Target Gram–Schmidt Procedure

- **Objective** Be able to apply the Gram–Schmidt Procedure to vectors from any inner product space in order to produce an orthonormal basis for the span of the vectors.

Student Learning Targets, and Objectives 2 of 2

Target Linear Functionals on Inner Product Spaces

- **Objective** Be able to apply Riesz Representation Theorem to "describe" a general function in $\mathcal{L}(V, \mathbb{F})$ as an inner product on V.

Target Calculating (Minimum) Distance to a Subspace

- **Objective** Be able to use projections in order to determine the minimum distance to a subspace.

Time-Target: 3×75-minute lectures.

Introduction :: Inner Products

So far, we have not talked about the length/size/norm of vectors (not even in $\mathbb{R}^n$); the familiar norm (the "2-norm" or "Euclidean norm") defined by

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}, \quad x \in \mathbb{R}^n$$

is *not linear (as a function of the components of x)* on $\mathbb{R}^n$ and thus does not fit in with the previous discussion of Linear Vector Spaces...

At this point we are ready to add the notion of length/size/norm of vectors, for vectors from all kinds of Vector Spaces, to our toolbox.

We start in $\mathbb{R}^n$, but quickly move to more general settings.

The Dot Product [Math 254]

Definition (Dot Product)

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

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n$$

Notation (Dot Product)

Note $\cdot : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ (two vectors in, one scalar out)

Properties (Dot Product)

- $x \cdot x \geq 0 \; \forall x \in \mathbb{R}^n$
- $x \cdot x = 0 \Leftrightarrow x = 0$
- $\forall y \in \mathbb{R}^n$ (fixed); $m_y : \mathbb{R}^n \mapsto \mathbb{R}$ defined by $m_y(x) = x \cdot y$ is linear.
- $x \cdot y = y \cdot x, \; \forall x, y \in \mathbb{R}^n$

From the Dot Product to the Inner Product

A map (like the dot product) which is linear once (any) one of the arguments is held fixed is sometimes referred to as being **bi-linear**.

In order to define a useful generalization of the dot product (which we will name an "inner product"), we first have to cover the complex case.

For $z = a + bi$, where $a, b \in \mathbb{R}$ ($z \in \mathbb{C}$):

- $|z| \stackrel{\text{def}}{=} \sqrt{a^2 + b^2}$
- $z^* = a - bi$
- $zz^* = z^*z = a^2 + b^2 = |z|^2$

With this in mind it is not a big leap to generalize the dot product to complex vectors as

$$\langle u,\, v\rangle = u_1 v_1^* + \cdots + u_n v_n^*, \quad \text{where } u, v \in \mathbb{C}^n$$

Inner Product

Definition (Inner Product)

An **inner product** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u,\, v\rangle \in \mathbb{F}$ and has the following properties:

Positivity:
$\langle v,\, v\rangle \geq 0 \; \forall v \in V$

Definiteness:
$\langle v,\, v\rangle = 0 \; \Leftrightarrow v = 0$

Additivity in the first argument:
$\langle u + v,\, w\rangle = \langle u,\, w\rangle + \langle v,\, w\rangle \; \forall u, v, w \in V$

Homogeneity (linear scaling) in the first argument‡:
$\langle \lambda u,\, v\rangle = \lambda \langle u,\, v\rangle \; \forall u, v \in V, \lambda \in \mathbb{F}$

Conjugate symmetry‡:
$\langle u,\, v\rangle = \langle v,\, u\rangle^* \; \forall u, v \in V$

Other properties follow from these...

‡ Note that with these definitions $\langle u,\, \lambda v\rangle = \lambda^* \langle u,\, v\rangle$; many physicists and some engineers prefer a definition with homogeneity in the second argument, so that $\langle u,\, \lambda v\rangle = \lambda \langle u,\, v\rangle$, and $\langle \lambda u,\, v\rangle = \lambda^* \langle u,\, v\rangle$. This Serves as Your Official Warning!!!

Inner Products :: Examples

Example (Inner Products)

- We have already introduced the Euclidean inner product on $\mathbb{F}^n$: $\langle w, z\rangle = w_1 z_1^* + \cdots + w_n z_n^*$
- If $c_1, \ldots, c_n$ are positive (and therefore real) numbers, then $\langle w, z\rangle = c_1 w_1 z_1^* + \cdots + c_n w_n z_n^*$ defines a **weighted** inner product on $\mathbb{F}^n$.
- Let $f, g \in C[-1, 1]$ (continuous on the interval $[-1, 1]$) be complex-valued functions, then we can define an inner product by

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

Inner Products :: Examples

Example (Inner Products)

- There are all kinds of interesting and useful inner products for real-valued polynomials $\mathcal{P}(\mathbb{R})$, *e.g.*

$$\langle f, g\rangle = \int_{-1}^{1} f(x) g(x) \, dx \qquad \text{[Legendre]}$$

$$\langle f, g\rangle = \int_{0}^{\infty} f(x) g(x) \, x^\alpha e^{-x} dx \qquad \text{[Laguerre]}$$

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x) g(x) \, e^{\frac{-x^2}{2}} dx \qquad \text{[Hermite]}$$

$$\langle f, g\rangle = \int_{-1}^{1} f(x) g(x) \frac{dx}{\sqrt{1-x^2}} \qquad \text{[Chebyshev]}$$

Among other things these polynomial inner products, and extensions eventually lead to Spherical Harmonics, Bessel Functions, Hankel Functions...

Inner Product Spaces

Definition (Inner Product Space)

An **inner product space** is a vector space V along with an inner product on V.

Note that a particular inner product "specializes" the vector space. In everything we did up to and including Eigenvalues and Eigenspaces was (maybe painfully?) general for all vector spaces.

If you are CS-object-oriented-inclined, you can think of vector spaces as base-classes with (virtual?) linear operators on them; and inner product spaces are the first level of derived classes.

Inner Product Spaces :: Notation, and Properties

With a slight abuse (or "overload"?) of notation we now let

Notation (V — Inner Product Space)

From now on, V denotes an inner product space over $\mathbb{F}$.

Theorem (Basic Properties of an Inner Product)

(a) *For $u \in V$ fixed, the function $\langle v, u\rangle : V \mapsto \mathbb{F}$ is a linear map.*

(b) $\langle 0, u\rangle = 0 \; \forall u \in V$

(c) $\langle u, 0\rangle = 0 \; \forall u \in V$

(d) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle \; \forall u, v, w \in V$

(e) $\langle u, \lambda v\rangle = \lambda^* \langle u, v\rangle \; \forall u, v \in V, \; \lambda \in \mathbb{F}$

The proof is straight-forward from definitions, and properties of complex numbers.

Norms (Inner Product ⇒ Norm)

The "inspiration" for inner products came from the dot-product in $\mathbb{R}^n$, which is tightly connected with the geometric-length / size / norm of a vector $v \in \mathbb{R}^n$.

Each inner product determines a norm:

Definition (Norm, $\|v\|$)

For $v \in V$, the **norm** of v, denoted $\|v\|$ is defined by

$$\|v\| = \sqrt{\langle v, v \rangle}$$

A Vector space with a norm is referred to as a **normed space**. Normed vector spaces are a superset of inner product spaces.

Basic Properties of the Norm (Norm $\not\Rightarrow$ Inner Product)

Theorem (Basic Properties of the Norm)

Suppose $v \in V$:

(a) $\|v\| = 0 \Leftrightarrow v = 0$

(b) $\|\lambda v\| = |\lambda| \, \|v\|, \forall \lambda \in \mathbb{F}$

Clearly, all functions induced by the inner product, $\|v\| = \sqrt{\langle v, v \rangle}$, will satisfy the above. (But the converse is not true).

"All inner products induce norms; but not all norms can be 'reverse engineered' to an inner product."

Again, the proof is by direct observation / computation.

Orthogonality

Definition (Orthogonal)

Two vectors $u, v \in V$ are **orthogonal** if $\langle u, v \rangle = 0$.

Sometimes we say that "u is orthogonal to v".

We use the notation $u \perp v$ to indicate orthogonality.

It is worth noting that this is very general, and now we can talk about *e.g.* orthogonal functions, and orthogonal polynomials.

In particular the Legendre, Laguarre, Hermite, and Chebyshev polyonomials are the ones that are orthogonal with respect to the inner products given on slide 10.

Orthogonality

Theorem (Orthogonality and 0)

(a) $0 \perp v, \forall v \in V$

(b) $0 \in V$ *is the only vector that is orthogonal to itself.*

Theorem (Pythagorean Theorem)

Suppose $u, v \in V : u \perp v$, then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof (Pythagorean Theorem)

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2 \end{aligned}$$

Orthogonal Decomposition

Let $u, (v \neq 0) \in V$, we can write u as a scalar multiple of v plus a vector $\perp v$: let $c \in \mathbb{F}$ —

$$\begin{aligned} u &= cv + (u - cv) \\ \langle u, v \rangle &= c\langle v, v \rangle + \underbrace{\langle u - cv, v \rangle}_{0} \\ \frac{\langle u, v \rangle}{\langle v, v \rangle} &= c \end{aligned}$$

Thus,

$$u = \underbrace{\frac{\langle u, v \rangle}{\langle v, v \rangle} v}_{u^{\parallel}} + \underbrace{\left(u - \frac{\langle u, v \rangle}{\langle v, v \rangle} v \right)}_{u^{\perp}}$$

Here, we have re-introduced some notation $(u^{\parallel}, u^{\perp})$ from [MATH 254].

We have already used this type of decomposition with a slightly different flavor... in [LIVEMATH#5B-4].

Orthogonal Decomposition

We summarize the previous argument:

Theorem (Orthogonal Decomposition)

Suppose $u, v \in V$, with $v \neq 0$. Let $c = \frac{\langle u, v \rangle}{\langle v, v \rangle}$, and $w = u - c\,v$; then

$$\langle w, v \rangle = 0, \quad \text{and} \quad u = cv + w$$

We will make use of this to show the next (major!) theorem.

The Cauchy-[Bunyakovsky]-Schwarz Inequality

Theorem (The Cauchy-[Bunyakovsky]-Schwarz Inequality)

Suppose $u, v \in V$, then

$$|\langle u, v \rangle| \leq \|v\| \, \|u\|.$$

The statement is an equality if and only if $u = kv$, $k \in \mathbb{F}$.

Pythagoras $\sim 570 - 495$ BC.
Augustin-Louis Cauchy, 21 August 1789 – 23 May 1857. (French)
⇒ proof for sums (1821).
Viktor Yakovlevich Bunyakovsky, 16 December 1804 – 12 December 1889. (Russian, Cauchy's graduate student)
⇒ proof for integrals (1859).
Karl Hermann Amandus Schwarz, 25 January 1843 – 30 November 1921. (German)
⇒ Modern proof (1888).

The Cauchy-[Bunyakovsky]-Schwarz Inequality

Proof (The Cauchy-[Bunyakovsky]-Schwarz Inequality)

If $v = 0$, then we have "$0 = 0$" and we're done.

Assume $v \neq 0$, and use the orthogonal decomposition

$$u = \frac{\langle u, v \rangle}{\langle v, v \rangle} v + w, \quad w \perp v.$$

By the [PYTHAGOREAN THEOREM]

$$\|u\|^2 = \left\| \frac{\langle u, v \rangle}{\langle v, v \rangle} v \right\|^2 + \|w\|^2 = \frac{|\langle u, v \rangle|^2}{\langle v, v \rangle} + \|w\|^2 \geq \frac{|\langle u, v \rangle|^2}{\langle v, v \rangle}$$

Multiply through by $\langle v, v \rangle = \|v\|^2$ and we get $\|u\|^2 \, \|v\|^2 \geq |\langle u, v \rangle|^2$; taking the square-root gives us the result $|\langle u, v \rangle| \leq \|u\| \, \|v\|$.

The proof reveals that the inequality is an equality if and only if $w = 0$; which means that $u = kv$, $k \in \mathbb{F}$

The Cauchy-[Bunyakovsky]-Schwarz Inequality :: Examples

Example (The Cauchy-[Bunyakovsky]-Schwarz Inequality)

- If $x, y \in \mathbb{R}^n$, then
$$|x_1 y_1 + \cdots + x_n y_n|^2 \leq \left(x_1^2 + \cdots + x_n^2\right)\left(y_1^2 + \cdots + y_n^2\right)$$
- Let $f, g \in C[-1, 1]$, then
$$\left| \int_{-1}^{1} f(x)g(x)\,dx \right|^2 \leq \left(\int_{-1}^{1} (f(x))^2\,dx \right) \left(\int_{-1}^{1} (g(x))^2\,dx \right)$$
- Let $f, g \in \mathcal{P}(\mathbb{R})$, then
$$\left| \int_{-1}^{1} f(x)g(x) \frac{dx}{\sqrt{1-x^2}} \right|^2 \leq \left(\int_{-1}^{1} (f(x))^2 \frac{dx}{\sqrt{1-x^2}} \right) \left(\int_{-1}^{1} (g(x))^2 \frac{dx}{\sqrt{1-x^2}} \right)$$

Triangle Inequality

Theorem (Triangle Inequality)

Suppose $u, v \in V$, then
$$\|u + v\| \leq \|u\| + \|v\|$$
This is an equality if and only if one of u, v is a non-negative multiple of the other.

Geometric Interpretation and Implication:

- The length of each side of a triangle is less than the sum of the lengths of the other two sides.
- The shortest path between two points is a line segment.

Triangle Inequality :: Visualization

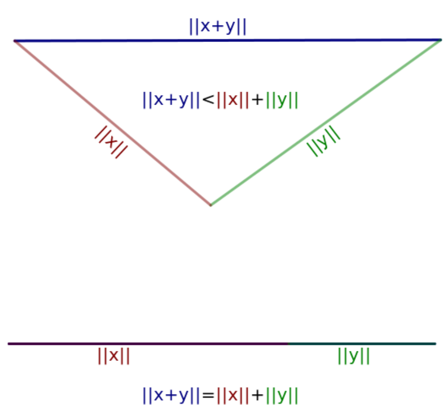


Figure: Illustration of the vectors involved in the triangle inequality.

Triangle Inequality

Proof (Triangle Inequality)

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v,\, u+v\rangle \\
&= \|u\|^2 + \|v\|^2 + \langle u,\, v\rangle + \langle v,\, u\rangle \\
&= \|u\|^2 + \|v\|^2 + \langle u,\, v\rangle + \langle u,\, v\rangle^* \\
&= \|u\|^2 + \|v\|^2 + 2\mathrm{Re}(\langle u,\, v\rangle) \\
\textcircled{1} &\leq \|u\|^2 + \|v\|^2 + 2|\langle u,\, v\rangle| \\
\textcircled{2} &\leq \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| \\
&= (\|u\| + \|v\|)^2 \\
\hline
\|u+v\| &\leq \|u\| + \|v\|
\end{aligned}$$

(2) follows from the CBS-inequality; in order to have an equality both (1) and (2) must be equalities $\Leftrightarrow$ one of u, v is a non-negative multiple of the other $\Leftrightarrow \langle u,\, v\rangle = \|u\|\,\|v\|$. (Minor details swept under the rug)

Parallelogram Equality

In every parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four side:

Theorem (Parallelogram Equality)

Suppose $u, v \in V$, then

$$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right)$$

Proof (Parallelogram Equality)

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v,\, u+v\rangle + \langle u-v,\, u-v\rangle \\ &= +\|u\|^2 + \|v\|^2 + \langle u,\, v\rangle + \langle v,\, u\rangle \\ &\quad +\|u\|^2 + \|v\|^2 - \langle u,\, v\rangle - \langle v,\, u\rangle \\ &= 2\left(\|u\|^2 + \|v\|^2\right)\end{aligned}$$

Parallelogram Equality :: Visualization

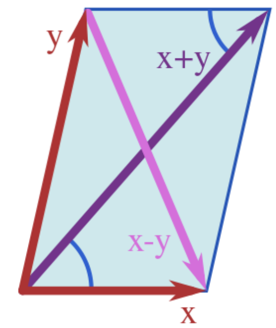


Figure: Illustration of the vectors involved in the parallelogram law.

$\langle\langle\langle$ Live Math $\rangle\rangle\rangle$

e.g. 6A-{**6**, **8**, 11, 12, 15, 19, 20, 22, 23}

Live Math :: Covid-19 Version 6A-6

6A-6: Suppose $u, v \in V$. Prove that

$$\langle u,\, v\rangle = 0 \quad \Leftrightarrow \quad \|u\| \le \|u + \alpha\, v\|,\ \forall \alpha \in \mathbb{F}$$

It is an if and only if , so we have 2 parts.

First suppose $\langle u,\, v\rangle = 0$.

Let $\alpha \in \mathbb{F}$. Since u and v are orthogonal, we can use [Pythagorean Theorem]:

$$\|u + \alpha\, v\| = \sqrt{\|u + \alpha v\|^2} = \sqrt{\|u\|^2 + \|\alpha\, v\|^2} \ge \sqrt{\|u\|^2} = \|u\|.$$

Live Math :: Covid-19 Version

Next, suppose $\|u\| \leq \|u + \alpha v\| \ \forall \alpha \in \mathbb{F}$

$$\begin{aligned}\|u\|^2 &\leq \|u+\alpha v\|^2 = \langle u+\alpha v,\, u+\alpha v\rangle \\ &= \langle u,\, u\rangle + \langle u,\, \alpha v\rangle + \langle \alpha v,\, u\rangle + \alpha\alpha^*\langle v,\, v\rangle \\ &= \|u\|^2 + 2\,\mathrm{Re}(\alpha^*\langle u,\, v\rangle) + |\alpha|^2\|v\|^2\end{aligned}$$

Since $\|u\|^2$ is non-negative, we must have

$$-2\,\mathrm{Re}(\alpha^*\langle u,\, v\rangle) \leq |\alpha|^2\|v\|^2$$

This is true $\forall \alpha$, let's write α in the form $-s\,\langle u,\, v\rangle$; then $\forall s > 0$:

$$\begin{aligned}2s\,|\langle u,\, v\rangle|^2 &\leq s^2\,|\langle u,\, v\rangle|^2\,\|v\|^2 \\ 2\,|\langle u,\, v\rangle|^2 &\leq s\,|\langle u,\, v\rangle|^2\,\|v\|^2\end{aligned}$$

The inequality holds whenever, $\|v\| = 0$ (which makes $\langle u,\, v\rangle = 0$), or consider $s < 2/\|v\|^2$ and conclude $\langle u,\, v\rangle = 0$. $\square$

Live Math :: Covid-19 Version

6A-8: Suppose $u, v \in V$, and $\|u\| = \|v\| = 1$ and $\langle u,\, v\rangle = 1$. Prove that $u = v$.

We have $|\langle u,\, v\rangle| = \|u\|\,\|v\|$; this means by [Cauchy-[Bunyakovsky]-Schwarz] that $u = \alpha v$, and

$$1 = \langle u,\, v\rangle = \langle \alpha v,\, v\rangle = \alpha\langle v,\, v\rangle = \alpha.$$

Orthonormality :: Building Blocks

Definition (Orthonormal)

- A list of vectors is called **orthonormal** if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list
- $u_1, \ldots, u_m \in U$ is orthonormal if (the Kronecker delta is back)

$$\langle u_j,\, u_k\rangle = \delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k\end{cases}$$

We care because:

Theorem (The Norm of an Orthonormal Linear Combination)

If $u_1, \ldots, u_m \in U$ is an orthonormal list of vectors in U, then $\forall a_1, \ldots, a_m \in \mathbb{F}$: $\|a_1 u_1 + \cdots + a_m u_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$.

Proof (The Norm of an Orthonormal Linear Combination)

$(m-1)$ applications of the Pythagorean Theorem.

Orthonormality :: Building Blocks

Theorem (An Orthonormal List is Linearly Independent)

Every orthonormal list of vectors is linearly independent.

Proof (An Orthonormal List is Linearly Independent)

$$0 = \|a_1 u_1 + \cdots + a_m u_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

forces $a_\ell \equiv 0$, thus $u_1, \ldots, u_m$ is linearly independent.

Definition (Orthonormal Basis)

An orthonormal basis of V is an orthonormal list of vectors in V that is also a basis of V.

Orthonormality :: Building Blocks

Theorem (An Orthonormal List of the Right Length is an Orthonormal Basis)

Every orthonormal list of vectors in V with length $\dim(V)$ is an orthonormal basis of V.

Proof (An Orthonormal List of the Right Length is an Orthonormal Basis)

By [An Orthonormal List is Linearly Independent] this list is linearly independent; and by [Linearly Independent List of length $\dim(V)$ is a Basis (Notes#2)] it is therefore a basis.

Example (An Orthonormal Basis of $\mathbb{F}^4$)

$$\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right),\ \left(\frac{1}{2},\frac{1}{2},-\frac{1}{2},-\frac{1}{2}\right),\ \left(\frac{1}{2},-\frac{1}{2},-\frac{1}{2},\frac{1}{2}\right),\ \left(-\frac{1}{2},\frac{1}{2},-\frac{1}{2},\frac{1}{2}\right)$$

Orthonormality :: Uses — Impact for Practical Computations

In general, given a basis $v_1, \ldots, v_n$ of V, and a vector $v \in V$, there are unique scalars $a_1, \ldots, a_n \in \mathbb{F}$, such that

$$v = a_1 v_1 + \cdots + a_n v_n$$

However, *computing* those coefficients typically requires serious work.

In the case of an orthonormal basis, this work is minimized to a single inner product for each scalar.

Theorem (Writing a Vector as a Linear Combination of Orthonormal Basis)

Suppose $v_1, \ldots, v_n$ is an orthonormal basis of V, and $v \in V$. Then

$$v = \langle v,\, v_1\rangle v_1 + \cdots + \langle v,\, v_n\rangle v_n$$

and

$$\|v\|^2 = |\langle v,\, v_1\rangle|^2 + \cdots + |\langle v,\, v_n\rangle|^2$$

Orthonormality

Proof (Writing a Vector as a Linear Combination of Orthonormal Basis)

Since $v_1, \ldots, v_n$ is an orthonormal basis of V, $\exists a_1, \ldots, a_n$:

$$\begin{aligned} v &= a_1 v_1 + \cdots + a_n v_n \\ \langle v,\, v_k\rangle &= \langle a_1 v_1 + \cdots + a_n v_n,\, v_k\rangle = a_k \end{aligned}$$

Clearly, orthonormal bases can greatly simplify some calculations. The next task is *constructing* them.

Figure: Jørgen Pedersen Gram (1850–1916)

Figure: Erhard Schmidt (1876–1959)

Gram–Schmidt Procedure

Theorem (Gram–Schmidt Procedure)

Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $u_1 = v_1 / \|v_1\|$. For $k = 2, \ldots, m$, define u_k by

$$u_k = \frac{v_k - \langle v_k,\, u_1\rangle u_1 - \cdots - \langle v_k,\, u_{k-1}\rangle u_{k-1}}{\|v_k - \langle v_k,\, u_1\rangle u_1 - \cdots - \langle v_k,\, u_{k-1}\rangle u_{k-1}\|}$$

The $u_1, \ldots, u_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(u_1, \ldots, u_k),\ \ k = 1, \ldots, m.$$

Note: This is *exactly* the same procedure you may (should) have seen for vectors in $\mathbb{R}^n$.

Gram–Schmidt Procedure

Comment (Detecting Linearly Dependent Vectors)

If we remove the assumption that $v_1, \ldots, v_m$ is linearly independent, the Gram–Schmidt procedure can be used to detect* linearly dependent vectors. If at any stage, the numerator (and therefore also the denominator) in the expression

$$u_k = \frac{v_k - \langle v_k,\, u_1\rangle u_1 - \cdots - \langle v_k,\, u_{k-1}\rangle u_{k-1}}{\| v_k - \langle v_k,\, u_1\rangle u_1 - \cdots - \langle v_k,\, u_{k-1}\rangle u_{k-1} \|}$$

becomes 0; then the vector $v_k \in \text{span}(v_1, \ldots, v_{k-1})$

* — at least in theory; in real life there may be some "issues", see [MATH 543].

Gram–Schmidt Procedure

Proof (Gram–Schmidt Procedure)

[PROOF-BY-(STRONG)-INDUCTION]

- $\ell = 1$: $\text{span}(v_1) = \text{span}(u_1)$, since v_1 is a positive multiple of u_1.
- Assume the theorem is true up to $(\ell - 1)$, where $1 < \ell \le m$. Since $v_1, \ldots, v_m$ is linearly independent $v_\ell \notin \text{span}(v_1, \ldots, v_{\ell-1}) = \text{span}(u_1, \ldots, u_{\ell-1})$; this means that the denominators in the theorem are non-zero, and the generated vectors have norm 1, $\|u_\ell\| = 1$
- Let $1 \le k < \ell$, then

$$\begin{aligned}\langle u_\ell,\, u_k\rangle &= \frac{\langle v_\ell,\, u_k\rangle - \langle v_\ell,\, u_1\rangle\langle u_1,\, u_k\rangle - \cdots - \langle v_\ell,\, u_{\ell-1}\rangle\langle u_{\ell-1},\, u_k\rangle}{\| v_\ell - \langle v_\ell,\, u_1\rangle u_1 - \cdots - \langle v_\ell,\, u_{\ell-1}\rangle u_{\ell-1}\|} \\ &= \frac{\langle v_\ell,\, u_k\rangle - \langle v_\ell,\, u_k\rangle}{\| v_\ell - \langle v_\ell,\, u_1\rangle u_1 - \cdots - \langle v_\ell,\, u_{\ell-1}\rangle u_{\ell-1}\|} = 0\end{aligned}$$

Gram–Schmidt Procedure

Proof (Gram–Schmidt Procedure)

[PROOF-BY-(STRONG)-INDUCTION]

- Therefore, $u_1, \ldots, u_\ell$ is an orthonormal list.

  From the expression for u_ℓ, we have that $v_\ell \in \text{span}(u_1, \ldots, u_\ell)$; and since $\text{span}(v_1, \ldots, v_{\ell-1}) = \text{span}(u_1, \ldots, u_{\ell-1})$

$$\text{span}(v_1, \ldots, v_\ell) \subset \text{span}(u_1, \ldots, u_\ell)$$

  Both lists are linearly independent; thus

$$\dim(\text{span}(v_1, \ldots, v_\ell)) = \dim(\text{span}(u_1, \ldots, u_\ell)) = \ell$$

  and hence

$$\text{span}(v_1, \ldots, v_\ell) = \text{span}(u_1, \ldots, u_\ell).\ \surd$$

Gram–Schmidt Procedure :: The Legendre Polynomials

"How hard can it be?!"

We outline the construction of an orthonormal basis of $\mathcal{P}_m(\mathbb{R})$, with inner product $\langle f,\, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$:

We start with the standard basis $\{1, x, x^2, x^3, \ldots\}$, start the process:

- $\|1\|^2 = \int_{-1}^{1} 1\,dx = 2.$ $\quad\rightsquigarrow\quad \boxed{u_0 = \sqrt{\tfrac{1}{2}}}$
- $x - \langle x,\, u_0\rangle u_0 = x - \left[\int_{-1}^{1} x \frac{1}{\sqrt{2}}\,dx\right] \frac{1}{\sqrt{2}} = x,$

  $\|x\|^2 = \int_{-1}^{1} x^2\,dx = \frac{2}{3}.$ $\quad\rightsquigarrow\quad \boxed{u_1 = \sqrt{\tfrac{3}{2}}\,x}$
- $x^2 - \langle x^2,\, u_0\rangle u_0 - \langle x^2,\, u_1\rangle u_1 = x^2 - \frac{1}{2}\int_{-1}^{1} x^2\,dx - 0 = x^2 - \frac{1}{2}\cdot\frac{2}{3} = x^2 - \frac{1}{3}.$

  $\left\|x^2 - \frac{1}{3}\right\|^2 = \int_{-1}^{1} \left(x^4 - \frac{2}{3}x^2 + \frac{1}{9}\right)\,dx = \frac{8}{45}.$ $\quad\rightsquigarrow\quad \boxed{u_2 = \sqrt{\tfrac{45}{8}}\left(x^2 - \tfrac{1}{3}\right)}$

Gram–Schmidt Procedure :: The Legendre Polynomials

- $x^3 - \langle x^3, u_0\rangle u_0 - \langle x^3, u_1\rangle u_1 - \langle x^3, u_2\rangle u_2 = x^3 - \langle x^3, u_1\rangle u_1$
  $= x^3 - \langle x^3, u_1\rangle u_1 = x^3 - \frac{\sqrt{6}}{5}\sqrt{\frac{3}{2}}x$
  $\left\|x^3 - \frac{3}{5}x\right\|^2 = \frac{8}{175}$. $\rightsquigarrow$ $u_3 = \sqrt{\frac{175}{8}}\left(x^3 - \frac{3}{5}x\right)$
- Yeah, it gets ugly fast! Usually, the Legendre Polynomials are listed in (one of) their orthogonal (but not ortho*normal*) form(s), *e.g.*:

$$1,\ x,\ \frac{1}{2}\left(3x^2-1\right),\ \frac{1}{2}\left(5x^3-3x\right),\ \frac{1}{8}\left(35x^4-30x^2+3\right),\ \ldots$$

- Using some software with symbolic calculation capabilities is useful in deriving these...

The Legendre Polynomials :: Comments

The Legendre polynomials are solutions to Legendre's differential equation:

$$\frac{d}{dx}\left[\left(1-x^2\right)\frac{dP_n(x)}{dx}\right] + n(n+1)P_n(x) = 0\,.$$

The orthogonality and completeness of these solutions is best seen from the viewpoint of Sturm–Liouville theory. [MATH 531]

There are many other examples of orthogonal functions/polynomials; of great interest are the trigonometric polynomials $\{\cos(n\theta), \sin(n\theta)\}$, or $\{e^{-in\theta}\}$; they form the basis for Fourier series expansions, which are the foundation for much of modern signal processing.

Let's return to our "safe?" linear algebra universe...

Existence of Orthonormal Basis

Theorem (Existence of Orthonormal Basis)

Every finite-dimensional inner product space has an orthonormal basis.

Proof (Existence of Orthonormal Basis)

Suppose V is finite-dimensional. Choose a basis of V. Apply the [GRAM–SCHMIDT PROCEDURE], producing an orthonormal list with length $\dim(V)$. By [AN ORTHONORMAL LIST OF THE RIGHT LENGTH IS AN ORTHONORMAL BASIS], this orthonormal list is an orthonormal basis of V.

Extending to an Orthonormal Basis

Theorem (Orthonormal List Extends to Orthonormal Basis)

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Proof (Orthonormal List Extends to Orthonormal Basis)

Suppose $u_1, \ldots, u_m$ is an orthonormal list of vectors in V. Then $u_1, \ldots, u_m$ is linearly independent [AN ORTHONORMAL LIST IS LINEARLY INDEPENDENT]. Hence this list can be extended to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ of V [LINEARLY INDEPENDENT LIST EXTENDS TO A BASIS (NOTES#2)]. Now apply the Gram–Schmidt Procedure to $u_1, \ldots, u_m, v_1, \ldots, v_n$, producing an orthonormal list $u_1, \ldots, u_m, w_1, \ldots, w_n$; here the formula given by the Gram–Schmidt Procedure leaves the first m vectors unchanged because they are already orthonormal. The list above is an orthonormal basis of V by [AN ORTHONORMAL LIST OF THE RIGHT LENGTH IS AN ORTHONORMAL BASIS].

Upper-triangular Matrix with respect to Orthonormal Basis

We have previously shown that if V is a finite-dimensional complex vector space, then for each operator on V there is a basis with respect to which the matrix of the operator is upper triangular. [Over $\mathbb{C}$, Every Operator has an Upper-Triangular Matrix (Notes#5)]

Theorem (Upper-triangular Matrix with respect to Orthonormal Basis)

Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Note: For real vector spaces, not all operator have an upper-triangular matrix with respect to some basis of V.

Upper-triangular Matrix with respect to Orthonormal Basis

Proof (Upper-triangular Matrix with respect to Orthonormal Basis)

Suppose T has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n$ of V. Thus $\text{span}(v_1, \ldots, v_k)$ is invariant under T for each $k \in \{1, \ldots, n\}$ [Conditions for Upper-Triangular Matrix (Notes#5)].

Apply the Gram–Schmidt Procedure to $v_1, \ldots, v_n$, producing an orthonormal basis $u_1, \ldots, u_n$ of V. Because

$$\text{span}(u_1, \ldots, u_k) = \text{span}(v_1, \ldots, v_k), \quad k \in \{1, \ldots, n\}$$

[Gram–Schmidt Procedure], we conclude that $\text{span}(u_1, \ldots, u_k)$ is invariant under T for each $k \in \{1, \ldots, n\}$. Thus, by [Conditions for Upper-Triangular Matrix (Notes#5)], T has an upper-triangular matrix with respect to the orthonormal basis $u_1, \ldots, u_n$.

Schur's Theorem ⇝ Schur's Matrix Decomposition

Theorem (Schur's Theorem)

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof (Schur's Theorem)

[Over $\mathbb{C}$, Every Operator has an Upper-Triangular Matrix (Notes#5)], [Upper-triangular Matrix with respect to Orthonormal Basis].

Schur's Theorem ⇝ Schur's Matrix Decomposition

Application (Schur Decomposition)

In computational linear algebra, the **Schur Decomposition** of a matrix $A \in \mathbb{C}^{n \times n}$ can be expressed as

$$A = QUQ^{-1}$$

where U is upper triangular, and Q unitary ($Q^{-1} = Q^*$).
(Every square matrix has a Schur decomposition)

Note: Not all mathematical results are useful in practical applications (they may require infinite-precision computing); however, the Schur Decomposition is stably computable in a finite precision environment.

Schur's Theorem ⇝ Schur's Matrix Decomposition

Application (Schur Decomposition :: Computation)

The Schur decomposition of a given matrix is numerically computed by the **QR algorithm** [Math 543] or its variants, *i.e.* the eigenvalues do not have to be pre-computed. (The eigenvalues show up as the diagonal entries of U).

The **QR algorithm** can be used to compute the roots of any given characteristic polynomial by finding the Schur decomposition of its companion matrix. — This is (one) numerically stable way to compute (good approximations of) eigenvalues of matrices.

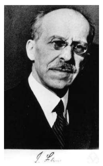


Figure: Issai Schur (1875–1941)

Linear Functionals on Inner Product Spaces

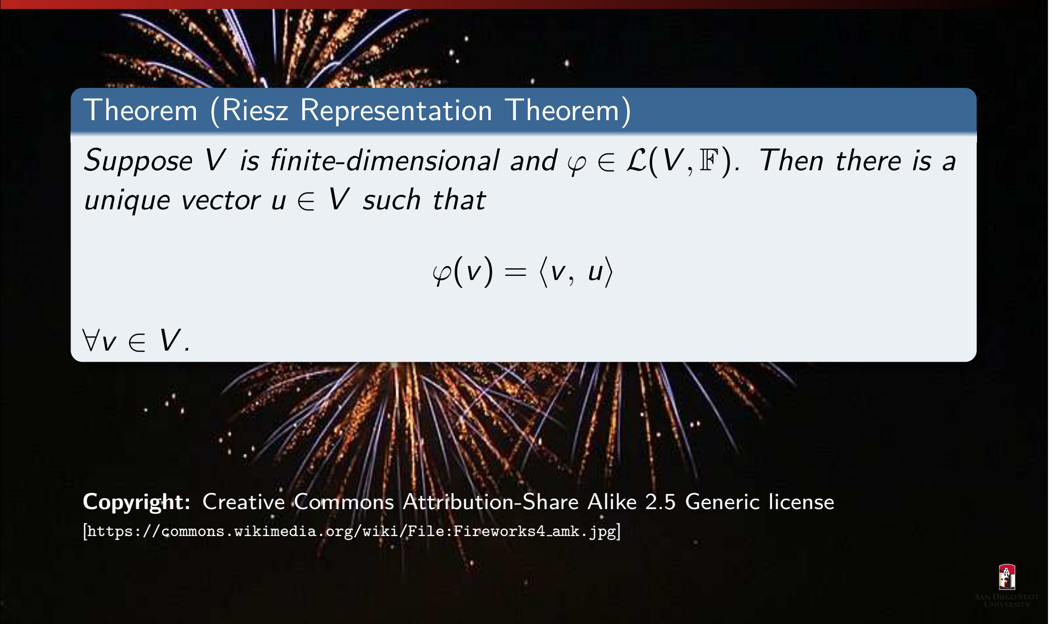
Definition (Linear Functional)

A **linear functional** on V is a linear map from $V \mapsto \mathbb{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbb{F})$.

Example (Linear Functional ⇝ Alternative Inner Product Form(?))

- $\varphi \in \mathcal{L}(\mathbb{F}^3, \mathbb{F})$ defined by $\varphi(z_1, z_2, z_3) = 2z_1 + 5z_2 + z_3$.
  Alternative form: $\varphi(z) = \langle z,\, u\rangle$, where $u = (2, 5, 1) \in \mathbb{F}^3$.
- $\varphi \in \mathcal{L}(\mathcal{P}_2(\mathbb{R}), \mathbb{R})$ defined by
$$\varphi(p) = \int_{-1}^{1} p(t) \underbrace{\cos(\pi t)}_{\notin \mathcal{P}_2(\mathbb{R})}\, dt$$
  It is not clear there there is an alternative form (in terms of the "Legendre" inner product on $\mathcal{P}_2(\mathbb{R})$), so that $\varphi(p) = \langle p,\, u\rangle$ for some $u \in \mathcal{P}_2(\mathbb{R})$.

Riesz Representation Theorem



Theorem (Riesz Representation Theorem)

Suppose V is finite-dimensional and $\varphi \in \mathcal{L}(V, \mathbb{F})$. Then there is a unique vector $u \in V$ such that
$$\varphi(v) = \langle v,\, u\rangle$$
$\forall v \in V$.

Riesz Representation Theorem :: Proof — Existence

Proof (Riesz Representation Theorem)

Existence:

- Let $u_1, \dots, u_n$ be an orthonormal basis of V; then
$$\begin{aligned}\varphi(v) &\overset{1}{=} \varphi(\langle v,\, u_1\rangle u_1 + \cdots + \langle v,\, u_n\rangle u_n)\\ &= \langle v,\, u_1\rangle\varphi(u_1) + \cdots + \langle v,\, u_n\rangle\varphi(u_n)\\ &= \langle v,\, \varphi(u_1)^* u_1\rangle + \cdots + \langle v,\, \varphi(u_n)^* u_n\rangle\\ &= \langle v,\, u\rangle\end{aligned}$$
  where $u = \varphi(u_1)^* u_1 + \cdots + \varphi(u_n)^* u_n \in V$.

$\overset{1}{=}$ [Writing a Vector as a Linear Combination of Orthonormal Basis].

Riesz Representation Theorem :: Proof — Uniqueness

Proof (Riesz Representation Theorem)

Uniqueness:

- Suppose $u_1, u_2 \in V$ such that

$$\varphi(v) = \langle v,\, u_1 \rangle = \langle v,\, u_2 \rangle$$

$\forall v \in V$; then

$$0 = \varphi(v) - \varphi(v) = \langle v,\, u_1 \rangle - \langle v,\, u_2 \rangle = \langle v,\, u_1 - u_2 \rangle$$

In particular $v = u_1 - u_2 \in V$, so that

$$\langle u_1 - u_2,\, u_1 - u_2 \rangle = 0$$

which forces $u_1 - u_2 = 0 \Leftrightarrow u_1 = u_2$. $\surd$

Riesz Representation Theorem :: Example

Consider, again, $\varphi \in \mathcal{L}(\mathcal{P}_2(\mathbb{R}), \mathbb{R})$ defined by

$$\varphi(p) = \int_{-1}^{1} p(t)\cos(\pi t)\, dt$$

[RRT] says we can find $u \in \mathcal{P}_2(\mathbb{R})$ so that

$$\int_{-1}^{1} p(t)\cos(\pi t)\, dt = \int_{-1}^{1} p(t)u(t)\, dt$$

$\forall p \in \mathcal{P}_2(\mathbb{R})$.

We use the expression $u = \varphi(u_1)^* u_1 + \ldots \varphi(u_n)^* u_n$, and the orthonormal basis

$$\mathfrak{B}_{\mathcal{P}_2(\mathbb{R})} = \left(\sqrt{\frac{1}{2}},\ \sqrt{\frac{3}{2}}x,\ \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right) \right)$$

Riesz Representation Theorem :: Example

"Nobody promised simple!"

We get

$$\begin{aligned} u(x) &= \left[\int_{-1}^{1} \sqrt{\frac{1}{2}}\cos(\pi t)\, dt\right]\sqrt{\frac{1}{2}} + \left[\int_{-1}^{1} \sqrt{\frac{3}{2}}\, t\cos(\pi t)\, dt\right]\sqrt{\frac{3}{2}}x \\ &\quad + \left[\int_{-1}^{1} \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right)\cos(\pi t)\, dt\right]\sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right) \\ &= \frac{1}{2}\left[\int_{-1}^{1} \cos(\pi t)\, dt\right] + \frac{3}{2}x\left[\int_{-1}^{1} t\cos(\pi t)\, dt\right] \\ &\quad + \frac{45}{8}\left(x^2 - \frac{1}{3}\right)\left[\int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)\cos(\pi t)\, dt\right] \\ &= 0 + 0 + \frac{45}{8}\left(x^2 - \frac{1}{3}\right)\frac{-4}{\pi^2} = \frac{-45}{2\pi^2}\left(x^2 - \frac{1}{3}\right) \end{aligned}$$

⟨⟨⟨ Live Math ⟩⟩⟩

e.g. 6B-{2, 4, **5**, **6**, 15}

Live Math :: Covid-19 Version — 6B-5

6B-5: On $\mathcal{P}_2(\mathbb{R})$, consider the inner product:

$$\langle p,\, q\rangle = \int_0^1 p(x)q(x)\,dx.$$

Apply the Gram–Schmidt procedure to $\{1, x, x^2\}$ to produce an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$.

$p_0(x) = 1$:

$$\langle 1,\, 1\rangle = \int_0^1 1^2, dx = x\big|_0^1 = 1 - 0 = 1.$$

Hence $\|1\| = 1$, and $u_0(x) = 1/\|1\| = 1$.

Live Math :: Covid-19 Version — 6B-5

$p_1(x) = x$: $\{u_0(x) = 1\}$

$$t_1(x) = p_1(x) - \langle p_1,\, u_0\rangle u_0(x),$$

$$\langle p_1,\, u_0\rangle = \int_0^1 x\,dx = \frac{x^2}{2}\bigg|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$$

$$t_1(x) = x - \frac{1}{2}1 = x - \frac{1}{2}$$

$$\|t_1\|^2 = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{3}\left(x - \frac{1}{2}\right)^3\bigg|_0^1 = \frac{1}{3}\left(\frac{1}{2}\right)^3 - \frac{1}{3}\left(\frac{-1}{2}\right)^3 = \frac{1}{3}\,\frac{1}{4}$$

$$u_1(x) = \frac{1}{\|t_1\|}t_1(x) = 2\sqrt{3}\left(x - \frac{1}{2}\right) = \sqrt{3}(2x-1)$$

Live Math :: Covid-19 Version — 6B-5

$p_2(x) = x^2$: $\{u_0(x) = 1,\ u_1(x) = \sqrt{3}(2x-1)\}$

$$t_2(x) = p_2(x) - \langle p_2,\, u_0\rangle u_0(x) - \langle p_2,\, u_1\rangle u_1(x),$$

$$\langle p_2,\, u_0\rangle = \int_0^1 x^2\,dx = \frac{1}{3}x^3\bigg|_0^1 = \frac{1}{3}$$

$$\langle p_2,\, u_1\rangle = \sqrt{3}\int_0^1 x^2(2x-1)\,dx = \sqrt{3}\left[\left(\frac{2}{4}x^4 - \frac{1}{3}x^3\right)\bigg|_0^1\right] = \sqrt{3}\left[\frac{3}{3\cdot 2} - \frac{2}{2\cdot 3}\right] = \frac{\sqrt{3}}{6}$$

$$t_2(x) = x^2 - \frac{1}{3}1 - \frac{\sqrt{3}}{6}\sqrt{3}(2x-1) = x^2 - \frac{1}{3} - \frac{1}{2}(2x-1) = x^2 - x + \frac{1}{6}$$

$$\begin{aligned}\|t_2\|^2 &= \int_0^1 \left(x^2 - x + \frac{1}{6}\right)^2 dx = \int_0^1 \left(x^4 - 2x^3 + \frac{4x^2}{3} - \frac{x}{3} + \frac{1}{36}\right) dx \\ &= \left(\frac{x^5}{5} - \frac{x^4}{2} + \frac{4x^3}{9} - \frac{x^2}{6} + \frac{x}{36}\right)\bigg|_0^1 = \frac{1}{5} - \frac{1}{2} + \frac{4}{9} - \frac{1}{6} + \frac{1}{36} = \frac{1}{180} = \frac{1}{2^2 3^2 5}\end{aligned}$$

$$u_2(x) = \frac{t_2(x)}{\|t_2(x)\|} = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right) = \sqrt{5}(6x^2 - 6x + 1)$$

Wasn't that fun?!? — $\{u_0(x) = 1,\ u_1(x) = \sqrt{3}(2x-1),\ u_2(x) = \sqrt{5}(6x^2 - 6x + 1)\}$.

Live Math :: Covid-19 Version — 6B-6

6B-6: Find an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$ (with inner product as in 6B-5) such that the differentiation operator $\mathcal{P}_2(\mathbb{R})$ has an upper-triangular matrix with respect to this basis.

First consider $\mathcal{M}(D, \{1, x, x^2\})$

Since, $D(1) = 0$, $D(x) = 1$, and $D(x^2) = 2x$:

$$\mathcal{M}(D, \{1, x, x^2\}) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Upper triangularity comes from the fact that

$$\mathrm{span}\,(1) \subset \mathrm{span}\,(1, x) \subset \mathrm{span}\,(1, x, x^2)$$

and each one of those spaces are invariant under D.

Live Math :: Covid-19 Version 6B-6

Now we are ready to consider $\mathcal{M}(D, \{1, \sqrt{3}(2x-1), \sqrt{5}(6x^2-6x+1)\})$

Clearly,

$$\text{span}\,(1) \subset \text{span}\left(1, \sqrt{3}(2x-1)\right) \subset \text{span}\left(1, \sqrt{3}(2x-1), \sqrt{5}(6x^2-6x+1)\right)$$

which means the orthonormal basis we found in 6B-5 satisfies our needs.

For extra giggles we find the matrix

$$D(1) = 0, \quad D(\sqrt{3}(2x-1)) = 2\sqrt{3}\cdot 1$$

$$D(\sqrt{5}(6x^2-6x+1)) = \sqrt{5}(12x-6) = \frac{6\sqrt{5}}{\sqrt{3}}\cdot\sqrt{3}(2x-1)$$

$$\mathcal{M}(D, \{1, \sqrt{3}(2x-1), \sqrt{5}(6x^2-6x+1)\}) = \begin{bmatrix} 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 2\sqrt{3}\sqrt{5} \\ 0 & 0 & 0 \end{bmatrix}$$

Orthogonal Complements

Definition (Orthogonal Complement, $U^\perp$)

If U is a subset of V, then the orthogonal complement of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u\rangle = 0\ \forall u \in U\}$$

Theorem (Properties of Orthogonal Complement)

(1) *If U is a subset (not a typo) of V, then $U^\perp$ is a subspace of V.*

(2) $\{0\}^\perp = V$

(3) $V^\perp = \{0\}$

(4) *If U is a subset of V, then $U^\perp \cap U \subset \{0\}$*

(5) *If U and W are subsets of V and $U \subset W$, then $W^\perp \subset U^\perp$*

Direct Sum of a Subspace and its Orthogonal Complement

Theorem (Direct Sum of a Subspace and its Orthogonal Complement)

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^\perp$$

Proof (Direct Sum of a Subspace and its Orthogonal Complement)

$V = U + U^\perp$:

Let $v \in V$; and $u_1, \ldots, u_m$ be an orthonormal basis of U; add a "clever 0" to v:

$$v = \underbrace{\langle v, u_1\rangle u_1 + \cdots + \langle v, u_m\rangle u_m}_{u} + \underbrace{v - \langle v, u_1\rangle u_1 + \cdots - \langle v, u_m\rangle u_m}_{w}$$

hence $u \in U$, and $\langle w, u_k\rangle = 0$, $k = 1, \ldots, m \Leftrightarrow w \perp \text{span}(u_1, \ldots, u_m)$, that is $w \in U^\perp$. We have written $v = u + w$, where $u \in U$ and $w \in U^\perp$. $\checkmark$

$V = U \oplus U^\perp$:

$V = U + U^\perp$ from above, and $U \cap U^\perp = \{0\}$ [Property (4)] $\Rightarrow V = U \oplus U^\perp$. $\checkmark$

Dimension

Theorem (Dimension of the Orthogonal Complement)

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim(U^\perp) = \dim(V) - \dim(U)$$

This follows directly from the previous theorem and:

Rewind ([A Sum is a Direct Sum ⇔ Dimensions Add Up (Notes#3.2)])

Suppose V is finite dimensional and $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim(U_1) + \cdots + \dim(U_m).$$

Complement-of-Complement

Theorem (The Orthogonal Complement of the Orthogonal Complement)

Suppose U is a finite-dimensional subspace of V. Then

$$(U^\perp)^\perp = U$$

Definition (Orthogonal Projection, $P_U(v)$)

Suppose U is a finite-dimensional subspace of V. The **orthogonal projection** of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined:
For $v \in V$, write $v = u + w$ where $u \in U$, $w \in U^\perp$; then

$$P_U(v) = u.$$

Since each $v \in V$ can be uniquely written in the form $v = u + w$ with $u \in U$, $w \in U^\perp$, the orthogonal projection is well defined.

Properties of the Orthogonal Projection

Theorem (Properties of the Orthogonal Projection)

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

(α) $P_U \in \mathcal{L}(V)$
(β) $P_U(u) = u, \forall u \in U$
(γ) $P_U(w) = 0, \forall w \in U^\perp$
(δ) $\text{range}(P_U) = U$
(ϵ) $\text{null}(P_U) = U^\perp$
(ζ) $(I - P_U)(v) = v - P_U(v) \in U^\perp$; $(I - P_U)$ *is the complementary projection*
(η) $(P_U)^2 = P_U$ [Live Math 5B-4]
(θ) $\|P_U(v)\| \le \|v\|$
(ι) *For every orthonormal basis $u_1, \dots, u_m$ of U:*

$$P_U(v) = \langle v,\, u_1\rangle u_1 + \cdots + \langle v,\, u_m\rangle u_m$$

The proofs are excellent training in application of the definitions and interactions of the various pieces involved.

Minimizing the Distance to a Subspace

In many applications we end up looking for the "best candidate" approximate solution to a problem; and it can usually be expressed in the linear algebra language as *"Given $v \in V$, find a point $u \in U$ such that $\|u - v\|$ is as small as possible:"*

Theorem (Minimizing the Distance to a Subspace)

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$, then

$$\|v - P_U(v)\| \le \|v - u\|$$

Equality holds if and only if $u = P_U(v)$.

That is, the "best candidate" approximate solution is given by the orthogonal projection onto the subspace.

This whole section "smells" like the foundation for numerical solutions of partial differential equations using the Finite Element Method (FEM)... and also has the obvious(?) application of model-fitting using linear-least-squares.

Minimizing the Distance to a Subspace

Proof (Minimizing the Distance to a Subspace)

$$\begin{aligned}\|v - P_U(v)\|^2 &\overset{1}{\le} \|\underbrace{v - P_U(v)}_{\in U^\perp}\|^2 + \|\underbrace{P_U(v) - u}_{\in U}\|^2 \\ &\overset{2}{=} \|v - P_U(v) + P_U(v) - u\|^2 \\ &= \|v - u\|^2\end{aligned}$$

$\overset{1}{\le}$ holds since $0 \le \|P_U(v) - u\|$; and
$\overset{2}{=}$ [Pythagorean Theorem].
We get an equality if and only if $P_U(v) - u = 0 \Leftrightarrow u = P_U(v)$. $\surd$

Example: sin(nx)

We consider the best Legendre-polynomial approximations of sin(nx) on the interval $[-\pi, \pi]$, and compare with the Taylor polynomials.

Nobody in their right mind would do this by hand!

With a little bit of Matlab-magic we can get it done.

Example: sin(nx) — Gram-Schmidt

```
clear x n
syms n x

%(Left and Right endpoints [a,b])
a = -pi;
b =  pi;

%(Symbolic, and Evaluated Innerproducts)
sip = @(u,v) int(u*v);
ip  = @(u,v) subs(sip(u,v),b)-subs(sip(u,v),a);

%(Starting Basis = powers of x)
p = [x.^(0:5)].';
u = 0*p;

%(Symbolic Gram-Schmidt)
for d=1:length(p)
  t = p(d);
  for prev = 1:(d-1)
    t = t - ip(t,u(prev)) * u(prev);
  end
  u(d) = t / sqrt(ip(t,t));
end
```


Example: sin(nx) — Gram-Schmidt Results

At this point we have our orthonormal basis:

$$\left\{ \frac{\sqrt{2}}{2\sqrt{\pi}},\quad \frac{\sqrt{6}\,x}{2\,\pi^{3/2}},\quad -\frac{\sqrt{10}\,\left(6\,\pi^2 - 18\,x^2\right)}{24\,\pi^{5/2}},\quad -\frac{\sqrt{14}\,x\,\left(6\,\pi^2 - 10\,x^2\right)}{8\,\pi^{7/2}}\right.$$

$$\left. \frac{\sqrt{2}\,\left(1680\,x^4 - 1440\,\pi^2\,x^2 + 144\,\pi^4\right)}{256\,\pi^{9/2}},\quad \frac{\sqrt{22}\,x\,\left(1260\,x^4 - 1400\,\pi^2\,x^2 + 300\,\pi^4\right)}{320\,\pi^{11/2}} \right\}$$

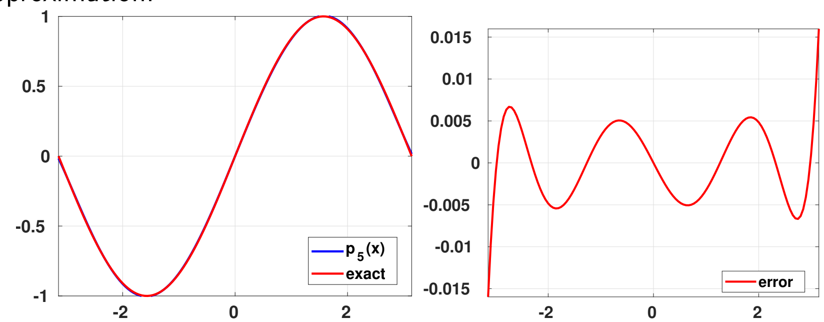
We set $n = 1$ in order to reproduce the results in the book

```
%(Function to Approximate)
v = @(x) sin(x);

%(Build Approximation)
f = 0*x;
for d=1:length(p)
  f = f + ip(v,u(d))*u(d);
end

%(Get the coefficients for the approximating polynomial)
poly_coeffs = coeffs(simplify(expand(expand(f))),'all').'
```


Example: sin(nx) — Best Polynomial Fit

$$\begin{aligned} p_5(x) &= \frac{693\,\left(\pi^4 - 105\,\pi^2 + 945\right)}{8\,\pi^{10}}\,x^5 - \frac{315\,\left(\pi^4 - 125\,\pi^2 + 1155\right)}{4\,\pi^8}\,x^3 + \frac{105\,\left(\pi^4 - 153\,\pi^2 + 1485\right)}{8\,\pi^6}\,x \\ &= 0.005643117976347\,x^5 - 0.155271410633430\,x^3 + 0.987862135574673\,x \end{aligned}$$

which, indeed, matches the result in the book; and it is a "pretty good" approximation:

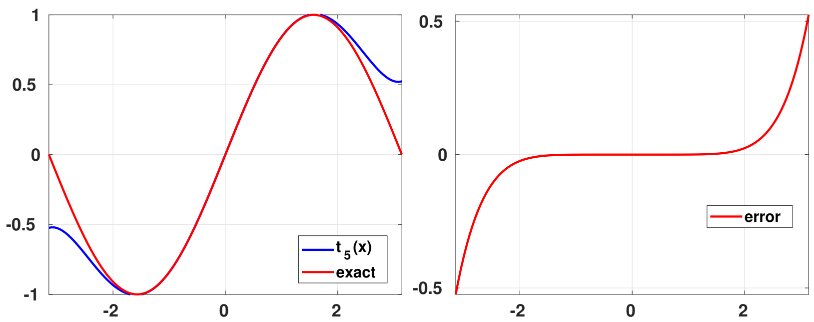


Example: $\sin(nx)$ — What about Taylor???

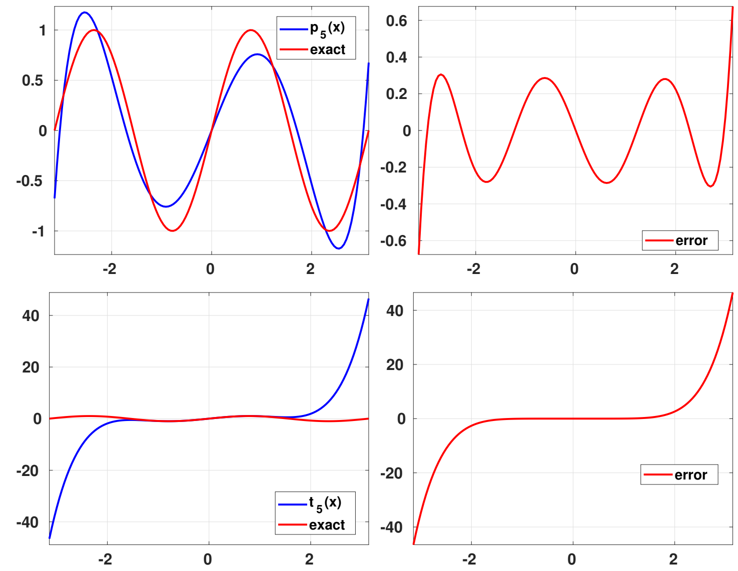
The fifth order Taylor polynomial is

$$t_5(x) = \frac{1}{120}x^5 - \frac{1}{6}x^3 + x$$

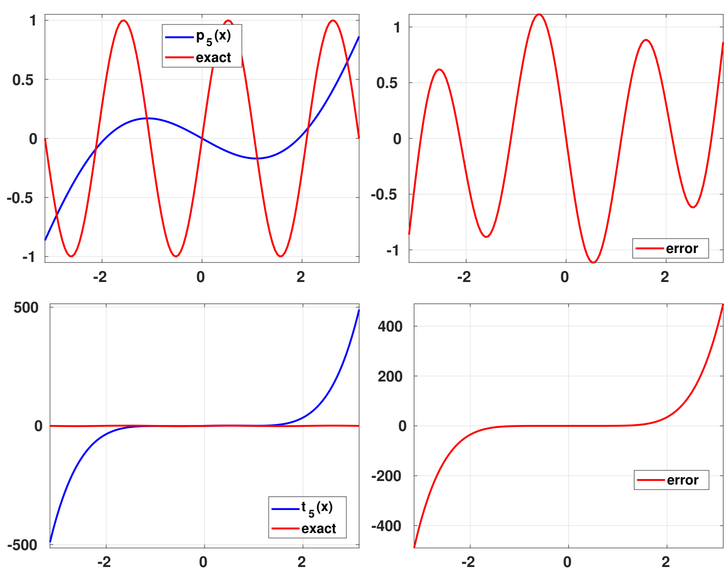
The approximation is not nearly as good:



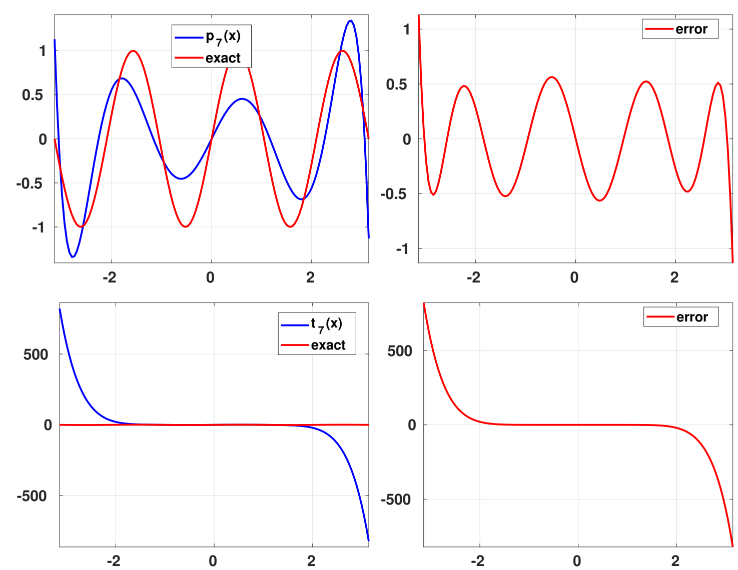
Example: Let's Have Fun — $\sin(2x)$, 5th Order Approximations



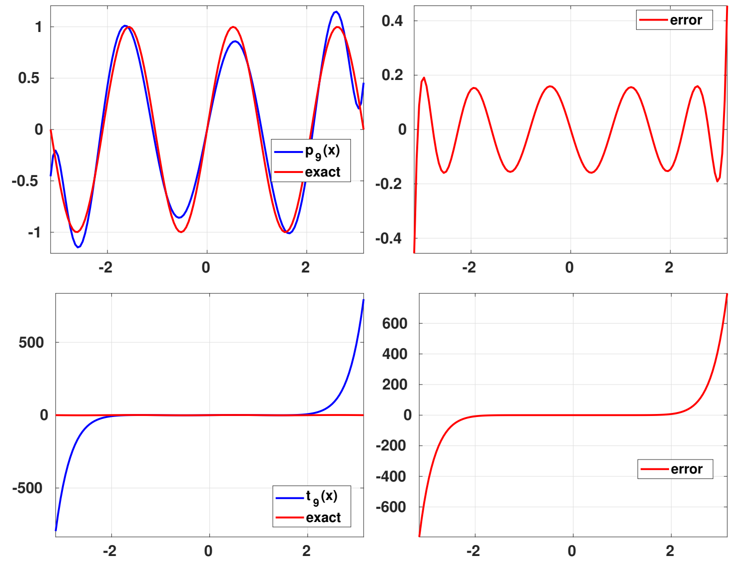
Example: Let's Have Fun — $\sin(3x)$, 5th Order Approximations



Example: Let's Have Fun — $\sin(3x)$, 7th Order Approximations



Example: Let's Have Fun — $\sin(3x)$, 9th Order Approximations



⟨⟨⟨ Live Math ⟩⟩⟩

e.g. 6C-{**2**, **3**}

Live Math :: Covid-19 Version 6C-2

6C-2: Suppose U is a finite-dimensional subspace of V. Prove that $U^\perp = \{0\}$ if and only if $U = V$.

From [Direct Sum of a Subspace and its Orthogonal Complement], we know that $V = U \oplus U^\perp$; which shows that $U^\perp = \{0\} \Leftrightarrow U = V$.

Comment

The statement does not hold on infinite dimensional inner product spaces. Both

- [Direct Sum of a Subspace and its Orthogonal Complement], and
- [The Orthogonal Complement of the Orthogonal Complement]

break without the finite-dimensional property.

As usual in mathematics, every word/property is there for a reason.

Live Math :: Covid-19 Version 6C-3

6C-3: Suppose U is a subspace of V with basis $u_1, \dots, u_m$ and

$$u_1, \dots, u_m, w_1, \dots, w_n$$

is a basis of V. Prove that if the Gram–Schmidt Procedure is applied to the basis of V above (from left-to-right), producing a list

$$f_1, \dots, f_m, g_1, \dots, g_n,$$

then $f_1, \dots, f_m$ is an orthonormal basis of U and $g_1, \dots, g_n$ is an orthonormal basis of $U^\perp$

Live Math :: Covid-19 Version 6C-3

Solution

We know that The Gram–Schmidt Procedure gives $f_1, \ldots, f_m$ so that $\mathrm{span}\,(f_1, \ldots, f_m) = \mathrm{span}\,(u_1, \ldots, u_m) = U$. Thus by [An Orthonormal List is Linearly Independent] $f_1, \ldots, f_m$ is an orthonormal basis of U.

Since $\langle g_k, f_j \rangle = 0$ ($\Leftrightarrow g_k \perp f_j$) $\forall k, j$;

$$g_k \perp \mathrm{span}\,(f_1, \ldots, f_m) = \mathrm{span}\,(u_1, \ldots, u_m) = U.$$

By definition, $g_k \in U^\perp$, $\forall k$. From [Dimension of the Orthogonal Complement], we have

$$\dim(U^\perp) = \dim(V) - \dim(U) = (m+n) - m = n$$

Thus $g_1, \ldots, g_n$ is an orthonormal basis of $U^\perp$.

Suggested Problems

6.A—1, 2, 4, 5, 6, 8, 11, 12, 15, 19, 20, 22, 23

6.B—1, 2, 3, 4, 6, 9, 15

6.C—5, 11

Assigned Homework HW#6, Due Date in Canvas/Gradescope

6.A—1, 2, 4, 5

6.B—1, 3, 9

6.C—5, 11

Note: Assignment problems are not official and subject to change until the first lecture on the chapter has been delivered (or virtually "scheduled.")

Upload homework to www.Gradescope.com

Other Versions of Riesz' Theorem Hilbert Space Representation

Let H be a Hilbert space, and let H^* denote its dual space, consisting of all continuous linear functionals from H into the field $\mathbb{R}$ or $\mathbb{C}$. If x is an element of H, then the function φ_x, for all y in H defined by: $\varphi_x(y) = \langle y, x \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of the Hilbert space, is an element of H^*. The Riesz representation theorem states that every element of H^* can be written uniquely in this form:

Theorem (Riesz–Fréchet Representation Theorem)

Let H be a Hilbert space and $\varphi \in H^*$. Then there exists $f \in H$ such that for any $x \in H$ $\varphi(x) = \langle f, x \rangle$. Moreover $\|f\|_H = \|\varphi\|_{H*}$

Statement source: https://en.wikipedia.org/wiki/Riesz_representation_theorem

Other Versions of Riesz' Theorem — Hausdorff Space Representation

Theorem (Riesz–Markov–Kakutani Representation Theorem)

Let X be a locally compact Hausdorff space. For any positive linear functional ψ on $C_c(X)$, there exists a unique regular Borel measure μ on X such that

$$\forall f \in C_c(X): \qquad \psi(f) = \int_X f(x)\, d\mu(x).$$

Statement source: https://en.wikipedia.org/wiki/Riesz–Markov–Kakutani_representation_theorem

Other Versions of Riesz' Theorem — Hausdorff Space Representation

Theorem (Riesz–Markov Representation Theorem)

Let X be a locally compact Hausdorff space. For any continuous linear functional ψ on $C_0(X)$, there exists a unique regular countably additive complex Borel measure μ on X such that

$$\forall f \in C_0(X): \qquad \psi(f) = \int_X f(x)\, d\mu(x).$$

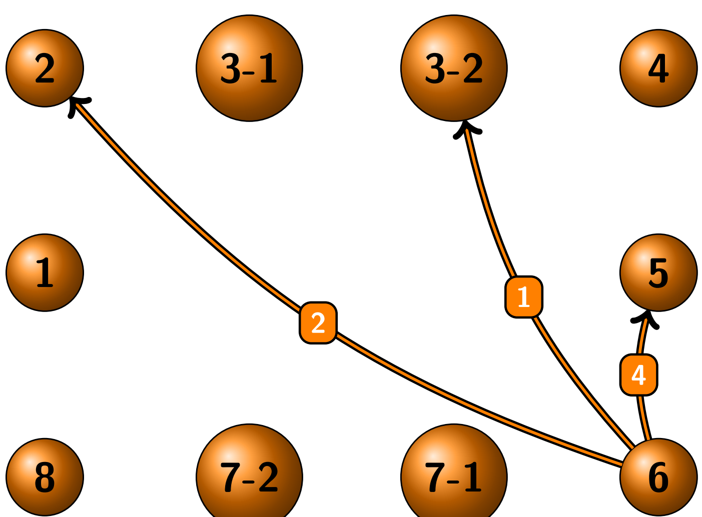
The norm of ψ as a linear functional is the total variation of μ, that is

$$\|\psi\| = |\mu|(X).$$

Finally, ψ is positive if and only if the measure μ is non-negative.

Statement source: https://en.wikipedia.org/wiki/Riesz–Markov–Kakutani_representation_theorem

Explicit References to Previous Theorems or Definitions (with count)



Explicit References to Previous Theorems or Definitions