Best approximation in the 2-norm

Elena Celledoni

Department of Mathematical Sciences, NTNU

september 26th 2012
A real vector space $V$ is a set with a 0 element and three operations:

- **Addition:** $x, y \in V$ then $x + y \in V$
- **Inverse:** for all $x \in V$ there is $-x \in V$ such that $x + (-x) = 0$
- **Scalar multiplication:** $\lambda \in \mathbb{R}$, $x \in V$ then $\lambda x \in V$

1. $x + y = y + x$
2. $(x + y) + z = x + (y + z)$
3. $0 + x = x + 0 = x$
4. $0x = 0$
5. $1x = x$
6. $(\lambda \mu) x = \lambda (\mu x)$
7. $\lambda (x + y) = \lambda x + \lambda y$
8. $(\lambda + \mu) x = \lambda x + \mu x$
Definition of inner product space
Let $V$ be a linear space over $\mathbb{R}$.
A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, is called an inner product on $V$ if:

1. $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$ for all $f, g, h$ in $V$;
2. $\langle \lambda f, g \rangle = \lambda \langle f, g \rangle$ for all $f, g$ in $V$ and $\lambda \in \mathbb{R}$;
3. $\langle f, g \rangle = \langle g, f \rangle$ for all $f, g$ in $V$;
4. $\langle f, f \rangle > 0$ if $f \neq 0$, $f \in V$. 

Example
The $n$-dimensional Euclidean space $\mathbb{R}^n$ is an inner product space $\langle u, b \rangle = \sum_{i=0}^{n} u_i v_i$, $u, v \in \mathbb{R}^n$ and $u = [u_1, \ldots, u_n]^T$, $v = [v_1, \ldots, v_n]^T$. 

Elena Celledoni
Definition of inner product space
Let $V$ be a linear space over $\mathbb{R}$.
A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, is called an inner product on $V$ if:

1. $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$ for all $f, g, h$ in $V$;
2. $\langle \lambda f, g \rangle = \lambda \langle f, g \rangle$ for all $f, g$ in $V$ and $\lambda \in \mathbb{R}$;
3. $\langle f, g \rangle = \langle g, f \rangle$ for all $f, g$ in $V$;
4. $\langle f, f \rangle > 0$ if $f \neq 0$, $f \in V$.

A linear space with an inner product is an inner product space.
Definition of inner product space
Let $V$ be a linear space over $\mathbb{R}$.
A function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$, is called an inner product on $V$ if:

1. $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$ for all $f, g, h$ in $V$;
2. $\langle \lambda f, g \rangle = \lambda \langle f, g \rangle$ for all $f, g$ in $V$ and $\lambda \in \mathbb{R}$;
3. $\langle f, g \rangle = \langle g, f \rangle$ for all $f, g$ in $V$;
4. $\langle f, f \rangle > 0$ if $f \neq 0$, $f \in V$.

A linear space with an inner product is an inner product space.

Example
The $n$-dimensional Euclidean space $\mathbb{R}^n$ is an inner product space

$$\langle u, b \rangle = \sum_{i=0}^{n} u_i v_i, \quad u, v \in \mathbb{R}^n$$

and $u = [u_1, \ldots, u_n]^T$, $v = [v_1, \ldots, v_n]^T$. 
Orthogonality
If in an inner product space \( V \) we have
\[
\langle f, g \rangle = 0
\]
then we say that \( f \) is orthogonal to \( g \).
Orthogonality
If in an inner product space $V$ we have

$$\langle f, g \rangle = 0$$

then we say that $f$ is orthogonal to $g$.

Induced norm
In an inner product space we can define the norm

$$\| f \| := \langle f, f \rangle^{\frac{1}{2}},$$

this is called norm induced by the inner product.
Orthogonality
If in an inner product space $V$ we have

$$\langle f, g \rangle = 0$$

then we say that $f$ is orthogonal to $g$.

**Induced norm**
In an inner product space we can define the norm

$$\|f\| := \langle f, f \rangle^{\frac{1}{2}},$$

this is called norm induced by the inner product.

**Theorem**  An inner product space $V$ over $\mathbb{R}$ with the induced norm defined above is a normed space.

**Proof.** Need to prove that the induced norm satisfies the axioms of norm, the proof makes use of the Cauchy-Schwarz inequality.
Lemma: Cauchy-Schwarz inequality

\[ |\langle f, g \rangle| \leq \|f\| \|g\|, \quad \forall f, g \in V. \]

Proof Similar to the one seen in chapter 2 for the inner product in \( \mathbb{R}^n \).
Lemma: Cauchy-Schwarz inequality

\[ |\langle f, g \rangle| \leq \|f\| \|g\|, \quad \forall f, g \in V. \]

Proof Similar to the one seen in chapter 2 for the inner product in \( \mathbb{R}^n \).

- Using the lemma we prove the triangular inequality for the induced norm:

\[ \|f + g\|^2 = \langle f + g, f + g \rangle = \|f\|^2 + 2\langle f, g \rangle + \|g\|^2 \]

from Cauchy-Schwarz we get

\[ \|f + g\|^2 \leq \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 = (\|f\| + \|g\|)^2 \]

so

\[ \|f + g\| \leq \|f\| + \|g\|. \]
Lemma: Cauchy-Schwarz inequality

\[ |\langle f, g \rangle| \leq \|f\| \|g\|, \quad \forall f, g \in V. \]

Proof Similar to the one seen in chapter 2 for the inner product in \(\mathbb{R}^n\).

- Using the lemma we prove the triangular inequality for the induced norm:

\[ \|f + g\|^2 = \langle f + g, f + g \rangle = \|f\|^2 + 2\langle f, g \rangle + \|g\|^2 \]

from Cauchy-Schwarz we get

\[ \|f + g\|^2 \leq \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 = (\|f\| + \|g\|)^2 \]

so

\[ \|f + g\| \leq \|f\| + \|g\|. \]

- The other two axioms of norm are directly proved using the definition of induced norm.
Example (9.3)

$C[a, b]$ (continuous functions defined on the closed interval $[a, b]$) is an inner product space with the inner product

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

and $w$ is a weight function defined, positive, continuous and integrable on $(a, b)$. The induced norm is

$$\|f\|_2 = \left(\int_a^b w(x)|f(x)|^2\,dx\right)^{1/2}.$$  

This norm is called “the 2-norm on $C[a, b]$”.
Space of square integrable functions: \( L^2 \)

We do not need \( f \) to be continuous in order for \( \|f\|_2 \) to be finite, it is sufficient that \( |f|^2w \) is integrable on \([a, b]\).
Space of square integrable functions: $L^2$

We do not need $f$ to be continuous in order for $\|f\|_2$ to be finite, it is sufficient that $|f|^2 w$ is integrable on $[a, b]$.

**Example**

$L_w^2(a, b) := \{ f : (a, b) \to \mathbb{R} \mid w(x)|f(x)|^2 \text{ is integrable on } (a, b) \}$. 

This is an inner product space with the inner product

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

and $w$ is a **weight function** defined, positive, continuous and integrable on $(a, b)$. The induced norm is

$$\|f\|_2 = \left( \int_a^b w(x)|f(x)|^2dx \right)^{\frac{1}{2}}.$$

This norm is called the $L^2$-norm. Note $C[a, b] \subset L_w^2(a, b)$. In the case $w \equiv 1$ then we write $L^2(a, b)$.
The problem of best approximation in the 2-norm can be formulated as follows:

**Best approximation problem in the 2-norm**

Given that \( f \in L^2_w(a, b) \), find \( p_n \in \Pi_n \) such that

\[
\| f - p_n \|_2 = \inf_{q \in \Pi_n} \| f - q \|_2
\]

such \( p_n \) is called a *polynomial of best approximation of degree \( n \) to the function \( f \) in the 2-norm on \( (a, b) \).
The problem of best approximation in the 2-norm can be formulated as follows:

**Best approximation problem in the 2-norm**

Given that \( f \in L_w^2(a, b) \), find \( p_n \in \Pi_n \) such that

\[
\| f - p_n \|_2 = \inf_{q \in \Pi_n} \| f - q \|_2
\]

such \( p_n \) is called a polynomial of best approximation of degree \( n \) to the function \( f \) in the 2-norm on \( (a, b) \).

**Theorem 9.2**

Let \( f \in L_w^2(a, b) \), there exists a unique polynomial \( p_n \in \Pi_n \) such that

\[
\| f - p_n \|_2 = \min_{q \in \Pi_n} \| f - q \|_2.
\]
Example: best approximation in $\| \cdot \|_2$ vs $\| \cdot \|_{\infty}$

Let $\epsilon > 0$ fixed

$$f(x) = 1 - e^{-\frac{x}{\epsilon}}, \quad x \in [0, 1]$$

find

$p_0^{2-norm}$ and $p_0^{\infty-norm}$
In general, to find $p_n(x) = c_0 + c_1 x + \cdots + c_n x^n$ best approximation in the two norm on $[0, 1]$: 

Solve $M c = b$ where $M$ is the Hilbert matrix: 

$$M_{j, k} = \int_0^1 x^k + j dx = \frac{1}{k + j + 1}, \quad k = 0, \ldots, n,$$

and 

$$b = \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix}, \quad b_j = \int_0^1 f(x) x^j dx.$$
In general, to find $p_n(x) = c_0 + c_1x + \cdots + c_nx^n$ best approximation in the two norm on $[0, 1]$: 

Solve

$$Mc = b$$

where $M$ is the Hilbert matrix:

$$M_{j,k} = \int_0^1 x^{k+j} \, dx = \frac{1}{k+j+1} \quad j, k = 0, \ldots, n,$$

and

$$b^T = \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix}, \quad b_j = \int_0^1 f(x)x^j \, dx.$$
Alternatively write \( p_n(x) = \gamma_0 \varphi_0 + \gamma_1 \varphi_1(x) + \cdots + \gamma_n \varphi_n(x) \) best approximation in the two norm on \([a, b]\\): 

and

\[
\Pi_n = \text{span}\{1, x, \ldots, x^n\} = \text{span}\{\varphi_0, \varphi_1 \ldots, \varphi_n\},
\]

and find \( \gamma_0, \ldots, \gamma_n \) minimizing

\[
E(\gamma_0, \ldots, \gamma_n) := \int_a^b |f(x) - \sum_{j=0}^{n} \gamma_j \varphi_j|^2.
\]
Alternatively write $p_n(x) = \gamma_0 \varphi_0 + \gamma_1 \varphi_1(x) + \cdots + \gamma_n \varphi_n(x)$: best approximation in the two norm on $[a, b]$:

and

$$\Pi_n = \text{span}\{1, x, \ldots, x^n\} = \text{span}\{\varphi_0, \varphi_1, \ldots, \varphi_n\},$$

and find $\gamma_0, \ldots, \gamma_n$ minimizing

$$E(\gamma_0, \ldots, \gamma_n) := \int_a^b |f(x) - \sum_{j=0}^n \gamma_j \varphi_j|^2.$$

This leads to

$$M \gamma = b$$

where $\gamma = [\gamma_0, \ldots, \gamma_n]^T$, $M$ is the matrix:

$$M_{j,k} = \langle \varphi_k, \varphi_j \rangle \quad j, k = 0, \ldots, n,$$

and

$$b = [b_0, \ldots, b_n]^T \quad b_j = \langle f, \varphi_j \rangle.$$

We’d like $M$ diagonal (i.e. $\langle \varphi_k, \varphi_j \rangle = \delta_{k,j}$).
Given a weight function $w$ on $(a, b)$ we say that the sequence of polynomials

$$
\varphi_j, \quad j = 0, 1, \ldots
$$

is a **system of orthogonal polynomials** on $(a, b)$ with respect to $w$, if each $\varphi_j$ is of exact degree $j$ and

$$
\langle \varphi_j, \varphi_k \rangle = \delta_{j,k}.
$$

To construct a system of orthogonal polynomials we can use a Gram-Schmidt process.
Gram-Schmidt orthogonalization

Let \( \varphi_0 \equiv 1 \) and assume \( \varphi_j \ j = 0, \ldots, n \ n \geq 0 \) are already orthogonal so

\[
\langle \varphi_k, \varphi_j \rangle = \int_a^b w(x) \varphi_k(x) \varphi_j(x) \, dx = 0, \quad k, j \in \{0, \ldots, n\}, \quad k \neq j
\]

consider

\[
q(x) = x^{n+1} - a_0 \varphi_0(x) - \cdots - a_n \varphi_n(x) \in \Pi_{n+1},
\]

and take

\[
a_j = \frac{\langle x^{n+1}, \varphi_j \rangle}{\langle \varphi_j, \varphi_j \rangle}, \quad j = 0, \ldots, n
\]

THEN
Let $\varphi_0 \equiv 1$ and assume $\varphi_j \; j = 0, \ldots, n \; n \geq 0$ are already orthogonal so

$$\langle \varphi_k, \varphi_j \rangle = \int_{a}^{b} w(x) \varphi_k(x) \varphi_j(x) \, dx = 0, \quad k, j \in \{0, \ldots, n\}, \quad k \neq j$$

consider

$$q(x) = x^{n+1} - a_0 \varphi_0(x) - \cdots - a_n \varphi_n(x) \in \Pi_{n+1},$$

and take

$$a_j = \frac{\langle x^{n+1}, \varphi_j \rangle}{\langle \varphi_j, \varphi_j \rangle}, \quad j = 0, \ldots, n$$

THEN

$$\langle q, \varphi_j \rangle = \langle x^{n+1}, \varphi_j \rangle - a_j \langle \varphi_j, \varphi_j \rangle = 0,$$
Gram-Schmidt orthogonalization

Let $\varphi_0 \equiv 1$ and assume $\varphi_j \; j = 0, \ldots, n \; n \geq 0$ are already orthogonal so

\[
\langle \varphi_k, \varphi_j \rangle = \int_a^b w(x) \varphi_k(x) \varphi_j(x) \, dx = 0, \quad k, j \in \{0, \ldots, n\}, \quad k \neq j
\]

consider

\[
q(x) = x^{n+1} - a_0 \varphi_0(x) - \cdots - a_n \varphi_n(x) \in \Pi_{n+1},
\]

and take

\[
a_j = \frac{\langle x^{n+1}, \varphi_j \rangle}{\langle \varphi_j, \varphi_j \rangle}, \quad j = 0, \ldots, n
\]

THEN

\[
\langle q, \varphi_j \rangle = \langle x^{n+1}, \varphi_j \rangle - a_j \langle \varphi_j, \varphi_j \rangle = 0,
\]

and we can take

\[
\varphi_{n+1} = \alpha q
\]

with $\alpha$ a scaling constant.
Example (9.6 Legendre polynomials)

\((a, b) = (-1, 1) \ w(x) \equiv 1:\)

\[ \begin{align*}
\varphi_0(x) & \equiv 1 \\
\varphi_1(x) & = x \\
\varphi_2(x) & = \frac{3}{2}x^2 - \frac{1}{2} \\
\varphi_2(x) & = \frac{5}{2}x^3 - \frac{3}{2}x
\end{align*} \]

Example (9.6 Chebishev polynomials)

\((a, b) = (-1, 1) \ w(x) \equiv (1 - x^2)^{-1/2} \)

\[ T_n(x) = \cos(n \arccos(x)), \quad n = 0, 1, \ldots. \]
**Theorem 9.3**

\[ p_n \in \Pi_n \text{ is such that} \]

\[ \| f - p_n \|_2 = \min_{q \in \Pi_n} \| f - q \|_2, \quad f \in L^2_w(a, b), \]

if and only if

\[ \langle f - p_n, q \rangle = 0, \quad \forall q \in \Pi_n \]

**Remark:** This theorem relates best approximation to orthogonality.
Example

On $(0, 1)$ and $w(x) \equiv 1$ on $(0, 1)$ by Gram-Schmidt orthogonalization we get

$$\varphi_0 = 1, \quad \varphi_1 = x - \frac{1}{2}, \quad \varphi_2(x) = x^2 - x + \frac{1}{6}.$$ 

So for

$$f(x) = 1 - e^{-\frac{x}{\epsilon}}, \quad x \in [0, 1]$$

the best approximation $p_3$ is $p_3(x) = \gamma_0 + \gamma_1 \varphi_1(x) + \gamma_2 \varphi_2(x)$
Example

On \((0, 1)\) and \(w(x) \equiv 1\) on \((0, 1)\) by **Gram-Schmidt orthogonalization** we get

\[
\varphi_0 = 1, \quad \varphi_1 = x - \frac{1}{2}, \quad \varphi_2(x) = x^2 - x + \frac{1}{6}.
\]

So for \(f(x) = 1 - e^{-\frac{x}{\epsilon}}, \quad x \in [0, 1]\)

the best approximation \(p_3\) is \(p_3(x) = \gamma_0 + \gamma_1 \varphi_1(x) + \gamma_2 \varphi_2(x)\)

\[
\gamma_0 = \int_0^1 \left(1 - e^{-\frac{x}{\epsilon}}\right) \, dx,
\]

\[
\gamma_1 = \frac{\int_0^1 \left(1 - e^{-\frac{x}{\epsilon}}\right)(x - \frac{1}{2}) \, dx}{\int_0^1 (x - \frac{1}{2})^2 \, dx},
\]

\[
\gamma_1 = \frac{\int_0^1 \left(1 - e^{-\frac{x}{\epsilon}}\right)(x^2 - x + \frac{1}{6}) \, dx}{\int_0^1 (x^2 - x + \frac{1}{6})^2 \, dx},
\]