Lesson 12

Or how to find a near solution of inconsistent systems

Chapter 6
Orthogonality and least squares

▸ Inner Product, Length and Orthogonality ▸ Orthogonal Sets ▸ Orthogonal Projections

▸ The Gram-Schmidt Process ▸ Least-Squares Problems ▸ Applications to Linear Models

▸ Inner Product Spaces ▸ Applications of Inner Product Spaces

The Gram-Schmidt Proces

Given a basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p\}$ for a subspace W of $\mathbb{R}^n$. Define:

$$
\begin{aligned}
&\boldsymbol{v}_1 = \boldsymbol{x}_1 \\
&\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x_2} \cdot \boldsymbol{v}_1}{\boldsymbol{v_1} \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 \\
&\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\boldsymbol{x_3} \cdot \boldsymbol{v}_1}{\boldsymbol{v_1} \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \frac{\boldsymbol{x_3} \cdot \boldsymbol{v}_2}{\boldsymbol{v_2} \cdot \boldsymbol{v}_2} \boldsymbol{v}_2 \\
&\vdots \\
&\boldsymbol{v}_p = \boldsymbol{x}_p - \frac{\boldsymbol{x_p} \cdot \boldsymbol{v}_1}{\boldsymbol{v_1} \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \frac{\boldsymbol{x_p} \cdot \boldsymbol{v}_2}{\boldsymbol{v_2} \cdot \boldsymbol{v}_2} \boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{x_p} \cdot \boldsymbol{v}_{p-1}}{\boldsymbol{v_{p-1}} \cdot \boldsymbol{v}_{p-1}} \boldsymbol{v}_{p-1}
\end{aligned}
$$

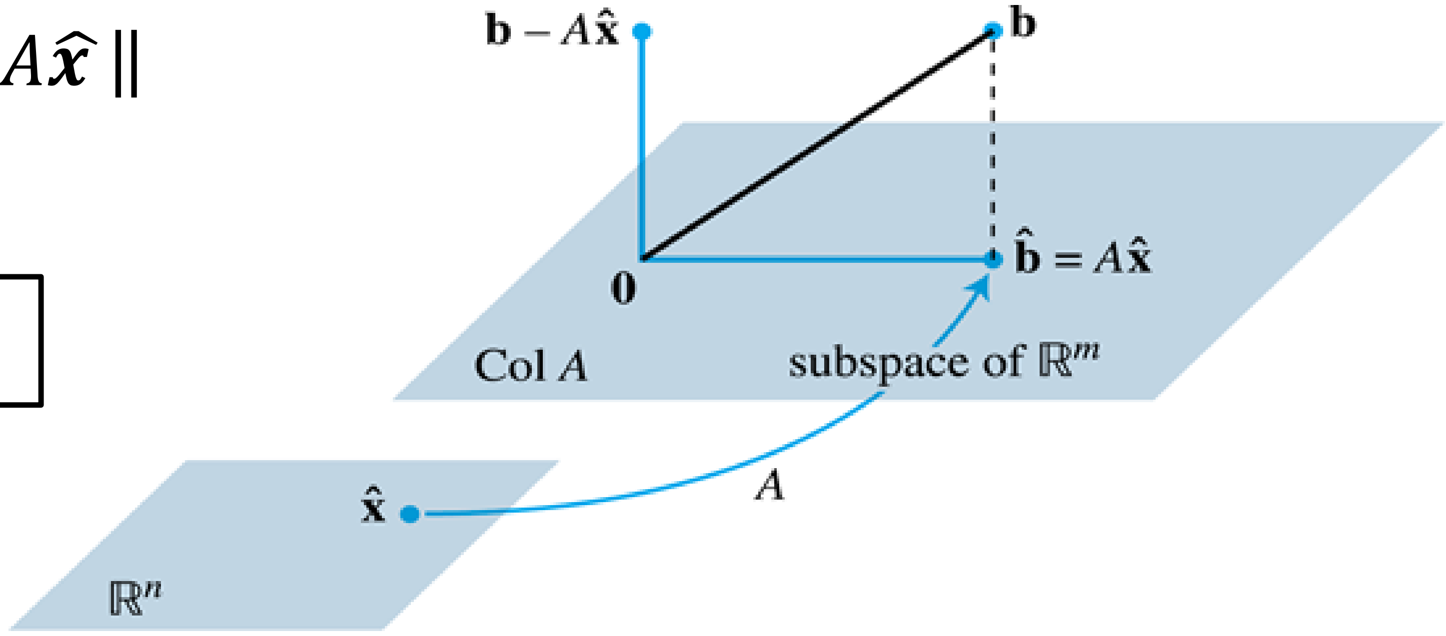
Then $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ is an orthogonal basis for W,

and $Span\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\} = Span\{\boldsymbol{v}_1, \cdots, \boldsymbol{v_k}\}$ for $k = 1, \cdots, p$

Least-squares problems

- Find an $\hat{\boldsymbol{x}}$ that makes $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|$ as small as possible

$$\boxed{A\hat{\boldsymbol{x}} = proj_{Col\ A}\boldsymbol{b} = \hat{\boldsymbol{b}}}$$



Theorem 6.13: The normal equations

The set of least-squares solutions of $A\boldsymbol{x} = \boldsymbol{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$

Least-Squares fitting of curves

Known (fitting) functions

Linear model = Linear in the unknown paratemers β_i

Fitting data to known functions: $y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x)$

Data points: $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

Determine the unknown parameters $\beta_0, \cdots, \beta_k$ that minimize the sum of square residuals:

$$\sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2, \qquad \hat{\boldsymbol{y}} = X\boldsymbol{\beta}$$

That is, find a least-squares solution of: $X\boldsymbol{\beta} = \boldsymbol{y}$

where:

Observation vector: $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Design matrix: $X = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_k(x_n) \end{bmatrix}$

Parameter vector: $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$

Residual vector: $\boldsymbol{\varepsilon} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

6.7 Inner Product Spaces

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle$$

Inner product (dot product): $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^{\boldsymbol{T}} \boldsymbol{v} = [u_1\ u_2\ \cdots\ u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i \in \mathbb{R}$$

- $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
- $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$
- $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v})$
- $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$ and $\boldsymbol{u} \cdot \boldsymbol{u} = 0 \Leftrightarrow \boldsymbol{u} = 0$
- The **norm** (or length) of a vector is defined as: $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + \cdots + v_n^2}$
- A vector $\boldsymbol{v}$ is **normalized** (unit vector) if: $\|\boldsymbol{v}\| = 1$
- The **distance** between two vectors is defined as:

$$dist(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{(u_1 - v_1)^{\boldsymbol{2}} + \cdots + (u_n - v_n)^{\boldsymbol{2}}}$$

- The **angle** θ between two vectors is given by: $\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\| \|\boldsymbol{v}\| \cos(\theta)$

Definition:

The **inner product** defined on a vector space V is a function that to any pair of vectors, **u** and **v** assigns a real number and fulfill the following axioms for all **u**, **v** and **w** in V and all scalars c:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$
2. $< \mathbf{u} + \mathbf{v}, \mathbf{w} > = < \mathbf{v}, \mathbf{w} > + < \mathbf{u}, \mathbf{w} >$
3. $< c\mathbf{u}, \mathbf{v} > = c < \mathbf{u}, \mathbf{v} >$
4. $< \mathbf{u}, \mathbf{u} > \geq 0$ and $< \mathbf{u}, \mathbf{u} > = 0 \iff \mathbf{u} = \mathbf{0}$

Ex 1 $\quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$: $\quad \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 4u_1v_1 + 5u_2v_2$ $\quad$ (Weighted dot-product)

Consider $\mathbb{R}^2$ and let an inner product be defined as

$$< \mathbf{u}, \mathbf{v} >= 2u_1v_1 + 3u_2v_2$$

Let two vectors be given as

$$\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

What is $< \mathbf{u}, \mathbf{v} >$?

Ex 2 $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n$: $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n)$; $t_0, t_1, \cdots, t_n \in \mathbb{R}$

Ex 3 $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_2$: $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + p(t_2)q(t_2)$; $t_0 = 0, t_1 = ½, t_2 = 1$

$\boldsymbol{p}(t) = 12t^2$; $\boldsymbol{q}(t) = 2t - 1$

$\boldsymbol{p}(t_0) = 0$; $\boldsymbol{q}(t_0) = -1$ $\quad \boldsymbol{p}(t_0)\boldsymbol{q}(t_0) = 0$

$\boldsymbol{p}(t_1) = 3$; $\boldsymbol{q}(t_1) = 0$ $\quad \boldsymbol{p}(t_1)\boldsymbol{q}(t_1) = 0$ $\qquad \langle \boldsymbol{p}, \boldsymbol{q} \rangle = 12$

$\boldsymbol{p}(t_2) = 12$; $\boldsymbol{q}(t_2) = 1$ $\quad \boldsymbol{p}(t_2)\boldsymbol{q}(t_2) = 12$

Diskussion:

Hvis t_2 i eksemplet på foregående slide, i stedet for at antage værdien 1, var ½, ville der gælde følgende:

- p og q er ortogonale for $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_2$, da $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = 0$

Vil ortogonaliteten også gælde for $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n$? Begrund svaret …

In a vector space where an inner product has been defined

The length / norm of a vector $\mathbf{v}$

$$\|\mathbf{v}\| = \sqrt{< \mathbf{v}, \mathbf{v} >}$$

The distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is

$$\|\mathbf{u} - \mathbf{v}\|$$

Two vectors are orthogonal if

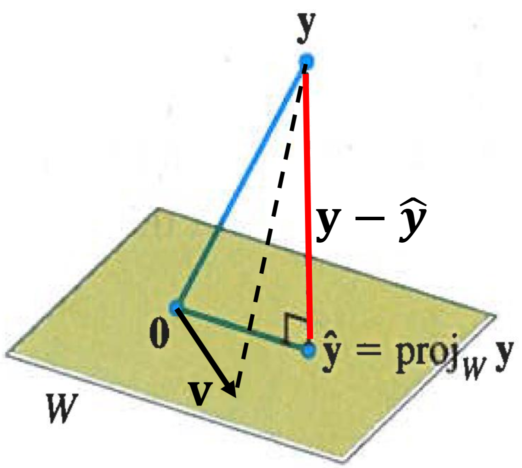
$$< \mathbf{u}, \mathbf{v} >= 0$$

Orthogonal Projection and Best Approximation

Let $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ be a orthogonal basis for a subspace W of vectorspace V, $\boldsymbol{y}$ any vector in V. Then:

$$\widehat{\boldsymbol{y}} = \frac{\langle \boldsymbol{y}, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 + \cdots + \frac{\langle \boldsymbol{y}, \boldsymbol{v}_p \rangle}{\langle \boldsymbol{v}_p, \boldsymbol{v}_p \rangle} \boldsymbol{v}_p = proj_W \boldsymbol{y}$$

is the orthogonal projection of $\boldsymbol{y}$ on W.



$\widehat{\boldsymbol{y}}$ is the closed point in W to $\boldsymbol{y}$ in the sense that:

$$\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| < \|\boldsymbol{y} - \boldsymbol{v}\|$$

for all $\boldsymbol{v}$ in W distinct from $\widehat{\boldsymbol{y}}$.

The Gram-Schmidt Proces

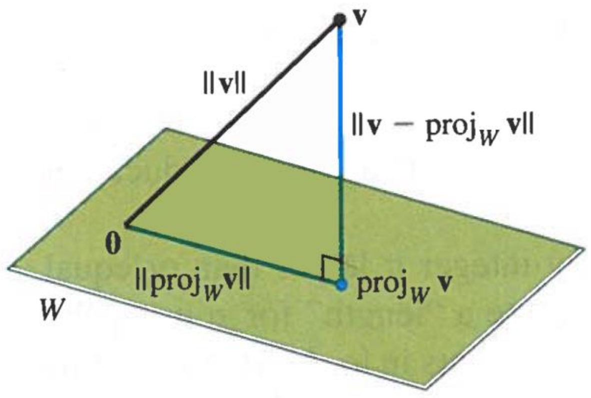
Given a basis $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p\}$ for a subspace W of vectorspace V. Define:

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{x}_1 \\
\boldsymbol{v}_2 &= \boldsymbol{x}_2 - \frac{\langle \boldsymbol{x}_2, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 \\
\boldsymbol{v}_3 &= \boldsymbol{x}_3 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 - \frac{\langle \boldsymbol{x}_3, \boldsymbol{v}_2 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle} \boldsymbol{v}_2 \\
&\vdots \\
\boldsymbol{v}_p &= \boldsymbol{x}_p - \frac{\langle \boldsymbol{x}_p, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1 - \frac{\langle \boldsymbol{x}_p, \boldsymbol{v}_2 \rangle}{\langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle} \boldsymbol{v}_2 - \cdots - \frac{\langle \boldsymbol{x}_p, \boldsymbol{v}_{p-1} \rangle}{\langle \boldsymbol{v}_{p-1}, \boldsymbol{v}_{p-1} \rangle} \boldsymbol{v}_{p-1}
\end{aligned}$$

Then $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p\}$ is an orthogonal basis for W and
Span$\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\} = Span\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ for $k = 1, \cdots, p$

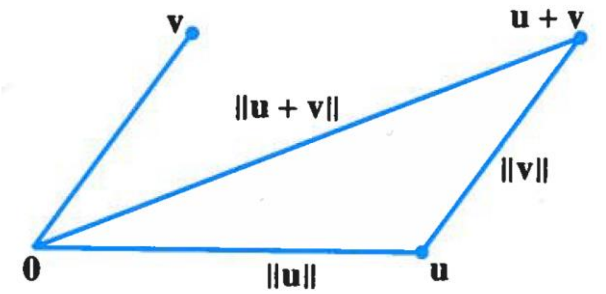
Therorem 6.16: The Cauchy-Schwartz Inequality

For all vectors $\boldsymbol{u}, \boldsymbol{v}$ in V: $\quad |\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$



Therorem 6.17: The Triangle Inequality

For all vectors $\boldsymbol{u}, \boldsymbol{v}$ in V: $\quad \|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$



Inner product for continuous functions

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_a^b p(t)q(t)dt$$

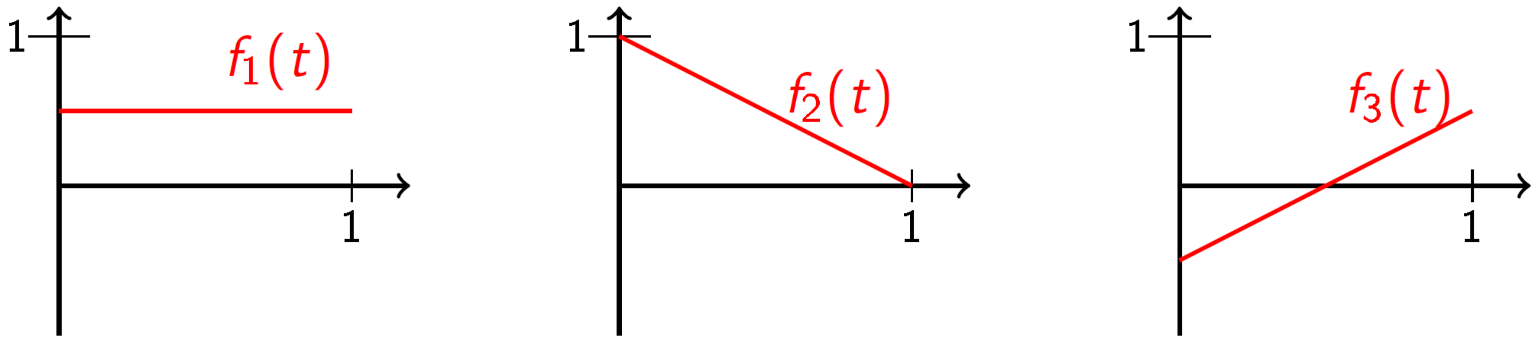
is an inner product of continuous functions on $[a, b]$

Funktionerne p og q er *ortogonale*, hvis integralet – evalueret mellem a og b – giver 0.

Let the inner product for continuous function defined on the interval $[0, 1]$ be given by

$$< f, g >= \int_0^1 f(t)g(t)dt$$

Three functions, $f_1(t)$, $f_2(t)$ and $f_3(t)$ are given as



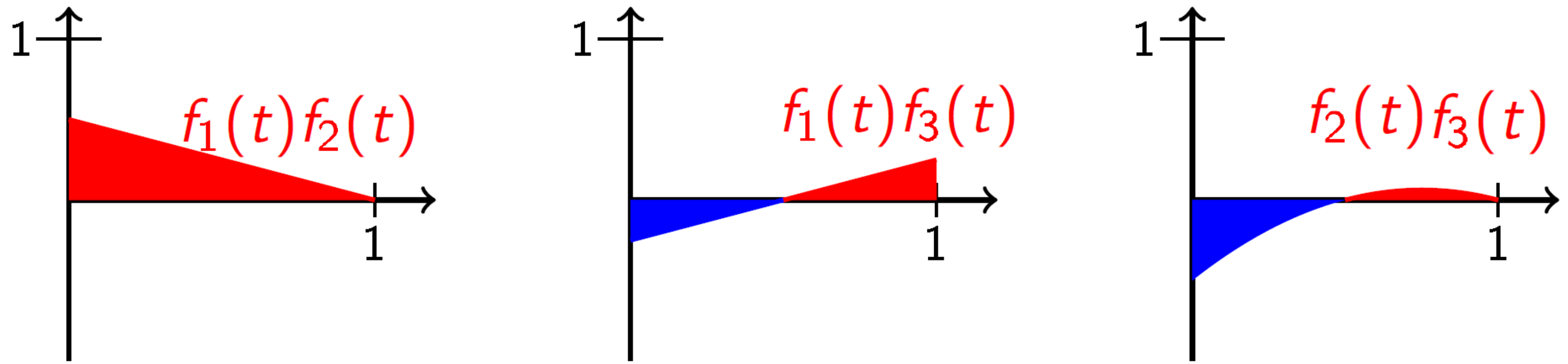
Which pair(s) of functions is orthogonal?

The functions are given by

$$f_1(t) = \frac{1}{2}, \qquad f_2(t) = 1 - t, \qquad f_3(t) = -\frac{1}{2} + t$$

and hence

$$f_1(t)f_2(t) = \frac{1}{2}(1-t), \quad f_1(t)f_3(t) = -\frac{1}{4} + \frac{1}{2}t, \quad f_2(t)f_3(t) = (1-t)(-\frac{1}{2}+t)$$



OK!

OPGAVE 1

Løs opgaverne:

4 og **6**

i eksamensopgavesættet fra 2012.

Eksamenssættet ligger tilgængeligt på Blackboard
under lektion 12.

6.8 Applications of Inner Product Spaces

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt$$

Weighted Least-Squares

n observations/measurements: $\boldsymbol{y} = (y_1, y_2, \cdots, y_n) \,\epsilon\, \mathbb{R}^n$

└→ some more accurate/reliable than others

The best approximation $\widehat{\boldsymbol{y}}$ in a subspace V of $\mathbb{R}^n$ (fx a curve) that minimize the sum of weighted squares of errors:

$$Weighted\ SS(E) = w_1^2(y_1 - \hat{y}_1)^2 + \cdots + w_n^2(y_n - \hat{y}_n)^2 = \|W\boldsymbol{y} - W\widehat{\boldsymbol{y}}\|^2$$

with the inner product:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = w_1^2 x_1 y_1 + \cdots + w_n^2 x_n y_n = (w_1 x_1)(w_1 y_1) + \cdots + (w_n x_n)(w_n y_n) = W\boldsymbol{x} \cdot W\boldsymbol{y}$$

and the weight diagonal matrix: $$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}$$

Weighted Least-Squares

Let $Col\ A = V$ then $\hat{\boldsymbol{y}} = A\hat{\boldsymbol{x}}$ and we seek a solution $\hat{\boldsymbol{x}}$ that minimize:

$$Weighted\ SS(E) = \|W\boldsymbol{y} - W\hat{\boldsymbol{y}}\|^2 = \|W\boldsymbol{y} - WA\hat{\boldsymbol{x}}\|^2$$

That is the ordinary least-squares solution to:

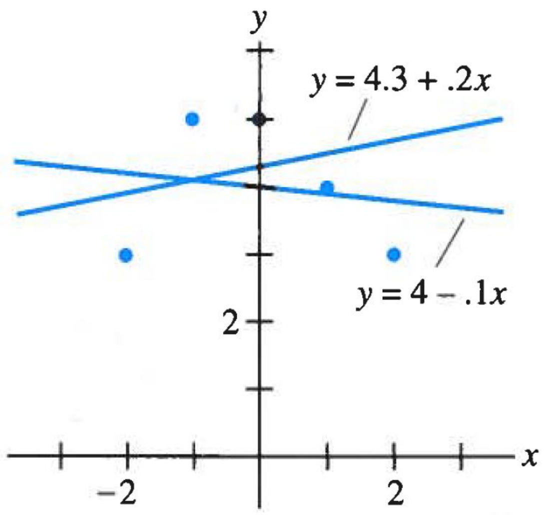
$$WA\boldsymbol{x} = W\boldsymbol{y}$$

With the normal equation:

$$(WA)^T WA\boldsymbol{x} = (WA)^T W\boldsymbol{y}$$

See Theorem 6.13

Weighted Least-Squares



5 measurements/data points: $(x, y) = \underbrace{(-2,3), (-1,5), (0,5)}, \underbrace{(1,4), (2,3)}$

Error:	Small	Large
Weight:	2	1

Best line: $y = \beta_0 + \beta_1 x$?

$$X = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \quad \boldsymbol{y} = \begin{bmatrix} 3 \\ 5 \\ 5 \\ 4 \\ 3 \end{bmatrix}; \quad W = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \implies WX = \begin{bmatrix} 2 & -4 \\ 2 & -2 \\ 2 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}; \quad W\boldsymbol{y} = \begin{bmatrix} 6 \\ 10 \\ 10 \\ 4 \\ 3 \end{bmatrix}$$

The normal equation: $(WX)^T WX\boldsymbol{\beta} = (WX)^T W\boldsymbol{y} \implies \begin{bmatrix} 14 & -9 \\ -9 & 25 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 59 \\ -34 \end{bmatrix} \implies \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 0.20 \end{bmatrix}$

$\implies$ **Weighted least-squares line: $\boldsymbol{y = 4.3 + 0.20x}$**

(Equal weighted least-squares line: $y = 4.0 - 0.10x$)

Fourier Series

Consider the set of trigonometric continuous functions in $C[0,2\pi]$:

$$\{1, \cos(t), \cos(2t), \cdots, \cos(nt), \sin(t), \sin(2t), \cdots, \sin(nt)\}$$

and the inner product:

$$\langle f, g\rangle = \int_0^{2\pi} f(t)g(t)dt$$

For all $m \neq n$:

$$\langle \cos(mt), \cos(nt)\rangle = \int_0^{2\pi} \cos(mt)\cos(nt)dt = 0$$

$$\langle \sin(mt), \sin(nt)\rangle = \int_0^{2\pi} \sin(mt)\sin(nt)dt = 0$$

$$\langle \cos(mt), \sin(nt)\rangle = \int_0^{2\pi} \cos(mt)\sin(nt)dt = 0$$

→ Orthogonal set of trigonometric continuous functions in $C[0,2\pi]$

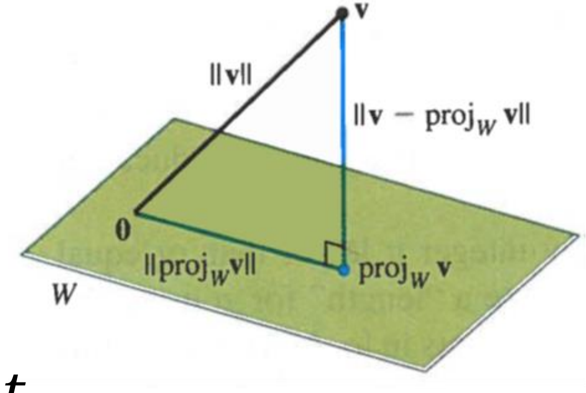
Fourier Series

Let $W = Span\{1, \cos(t), \cos(2t), \cdots, \cos(nt), \sin(t), \sin(2t), \cdots, \sin(nt)\}$
and $f \in C[0,2\pi]$. The projection of f on W:

$$\hat{f}(t) = proj_W f = \frac{a_0}{2} \cdot 1 + a_1\cos(t) + a_2\cos(2t) + \cdots a_n\cos(nt) + b_1\sin(t) + b_2\sin(2t) + \cdots + b_n\sin(nt)$$

Fourier series of order n

where



$$\frac{a_0}{2} = \frac{\langle f, 1\rangle}{\langle 1,1\rangle} = \frac{1}{2\pi}\int_0^{2\pi} f(t) \cdot 1dt$$

Fourier coefficients = Coordinates in the orthogonal trigonometric (harmonic/frequency) space W

$$a_k = \frac{\langle f, \cos(kt)\rangle}{\langle \cos(kt), \cos(kt)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(kt)dt$$

$$b_k = \frac{\langle f, \sin(kt)\rangle}{\langle \sin(kt), \sin(kt)\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(kt)dt$$

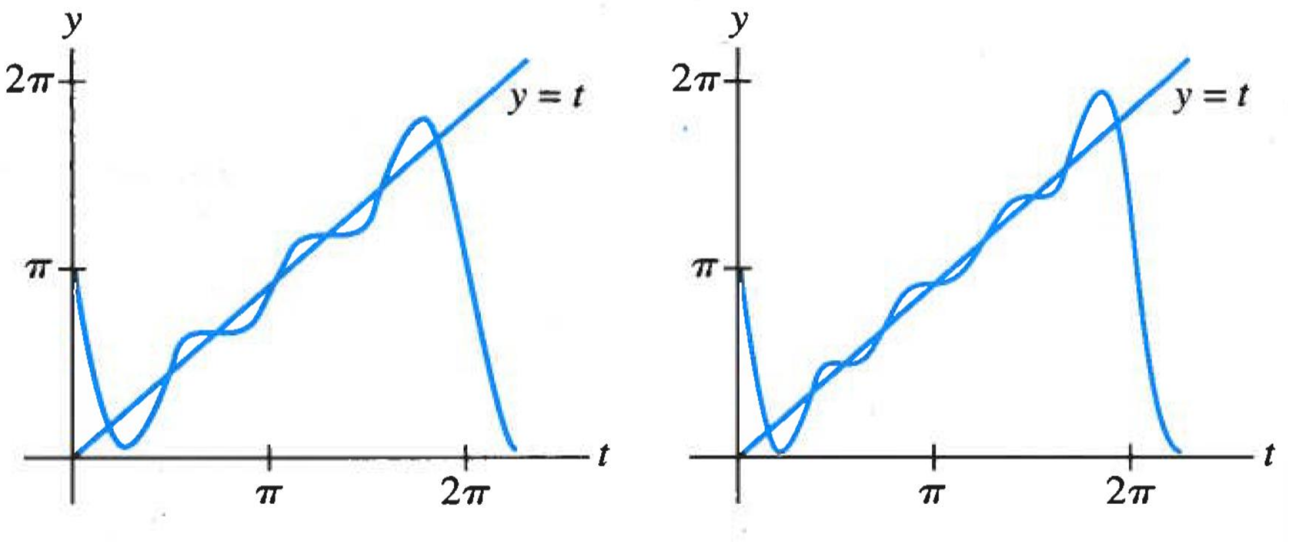
Fourier Series: $f(t) = t; \quad 0 \leq t \leq 2\pi$

$$\left.\begin{aligned} \frac{a_0}{2} &= \frac{1}{2\pi}\int_0^{2\pi} t\,dt = \pi \\ a_k &= \frac{1}{\pi}\int_0^{2\pi} t \cdot \cos(kt)dt = 0 \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} t \cdot \sin(kt)dt = -\frac{2}{k} \end{aligned}\right\} \Longrightarrow \hat{f}(t) = \pi - 2\sin(t) - \sin(2t) - \frac{2}{3}\sin(3t) - \cdots - \frac{2}{n}\sin(nt)$$

Mean square error:

$$\left\|f(t) - \hat{f}(t)\right\|^2$$

$$= \int_0^{2\pi}\left(f(t) - \hat{f}(t)\right)^2 dt \xrightarrow[n\to\infty]{} 0$$



(a) Third order (b) Fourth order

OPGAVE 2

For Fourier-serien på foregående slide, som approximerer funktionen $f(t) = t$:

- Bestem Fourier-koefficienterne for orden 3, 5 og 7
- Plot $f(t)$ samt de tre Fourier-serier i intervallet $[0; 2\pi]$
- Beregn kvadratfejlen for hver af de tre serier. Er tallene forventelige sammenlignet med hinanden?

Todays words and concepts

Weighted least-squares

Inner product

The Cauchy-Schwartz Inequality

Fourier Series

Trigonometric space

The Triangle Inequality

Inner product spaces

Harmonic projections