MATH 311

Topics in Applied Mathematics I

Lecture 26:
Least squares problems (continued).
Orthogonal bases.
The Gram-Schmidt orthogonalization process.

Least squares solution

System of linear equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \iff A\mathbf{x} = \mathbf{b}$$

For any $\mathbf{x} \in \mathbb{R}^n$ define a **residual** $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$.

The **least squares solution x** to the system is the one that minimizes $\|r(\mathbf{x})\|$ (or, equivalently, $\|r(\mathbf{x})\|^2$).

Theorem A vector $\hat{\mathbf{x}}$ is a least squares solution of the system $A\mathbf{x} = \mathbf{b}$ if and only if it is a solution of the associated **normal system** $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}.}$

Problem. Find the constant function that is the least squares fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c \implies \begin{cases} c = 1 \\ c = 0 \\ c = 1 \\ c = 2 \end{cases} \implies \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$(1, 1, 1, 1) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = (1, 1, 1, 1) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$c = \frac{1}{4}(1 + 0 + 1 + 2) = 1$ (mean arithmetic value)

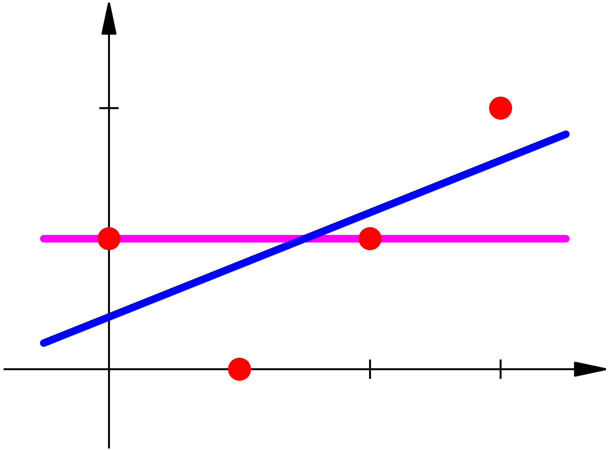
Problem. Find the linear polynomial that is the least squares fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2 x \implies \begin{cases} c_1 = 1 \\ c_1 + c_2 = 0 \\ c_1 + 2c_2 = 1 \\ c_1 + 3c_2 = 2 \end{cases} \implies \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} \iff \begin{cases} c_1 = 0.4 \\ c_2 = 0.4 \end{cases}$$



Problem. Find the quadratic polynomial that is the least squares fit to the following data

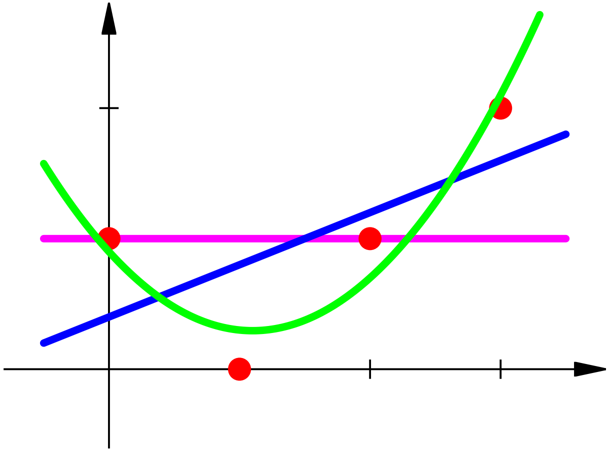
x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2 x + c_3 x^2$$

$$\implies \begin{cases} c_1 = 1 \\ c_1 + c_2 + c_3 = 0 \\ c_1 + 2c_2 + 4c_3 = 1 \\ c_1 + 3c_2 + 9c_3 = 2 \end{cases} \implies \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \\ 22 \end{pmatrix} \iff \begin{cases} c_1 = 0.9 \\ c_2 = -1.1 \\ c_3 = 0.5 \end{cases}$$



Orthogonal sets

Let $\langle\cdot,\cdot\rangle$ denote the scalar product in $\mathbb{R}^n$.

Definition. Nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$ form an **orthogonal set** if they are orthogonal to each other: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $i \neq j$.

If, in addition, all vectors are of unit length, $\|\mathbf{v}_i\| = 1$, then $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ is called an **orthonormal set**.

Example. The standard basis $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, ..., $\mathbf{e}_n = (0, 0, 0, \dots, 1)$. It is an orthonormal set.

Orthonormal bases

Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis for $\mathbb{R}^n$ (i.e., it is a basis and an orthonormal set).

Theorem Let $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n$, where $x_i, y_j \in \mathbb{R}$. Then

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$,

(ii) $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

Proof: (ii) follows from (i) when $\mathbf{y} = \mathbf{x}$.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left\langle \sum_{i=1}^n x_i\mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j \right\rangle = \sum_{i=1}^n x_i \left\langle \mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j \right\rangle$$

$$= \sum_{i=1}^n \sum_{j=1}^n x_iy_j\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{i=1}^n x_iy_i.$$

Suppose V is a subspace of $\mathbb{R}^n$. Let $\mathbf{p}$ be the orthogonal projection of a vector $\mathbf{x} \in \mathbb{R}^n$ onto V.

If V is a one-dimensional subspace spanned by a vector $\mathbf{v}$ then $\mathbf{p} = \dfrac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}$.

If V admits an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ then

$$\mathbf{p} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_k \rangle}{\langle \mathbf{v}_k, \mathbf{v}_k \rangle}\mathbf{v}_k.$$

Indeed, $\langle \mathbf{p}, \mathbf{v}_i \rangle = \sum_{j=1}^{k} \dfrac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}\langle \mathbf{v}_j, \mathbf{v}_i \rangle = \dfrac{\langle \mathbf{x}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}\langle \mathbf{v}_i, \mathbf{v}_i \rangle = \langle \mathbf{x}, \mathbf{v}_i \rangle$

$\implies \langle \mathbf{x}-\mathbf{p}, \mathbf{v}_i \rangle = 0 \implies \mathbf{x}-\mathbf{p} \perp \mathbf{v}_i \implies \mathbf{x}-\mathbf{p} \perp V.$

Coordinates relative to an orthogonal basis

Theorem If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for $\mathbb{R}^n$, then

$$\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n$$

for any vector $\mathbf{x} \in \mathbb{R}^n$.

Corollary If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis for $\mathbb{R}^n$, then

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{x}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n$$

for any vector $\mathbf{x} \in \mathbb{R}^n$.

The Gram-Schmidt orthogonalization process

Let V be a subspace of $\mathbb{R}^n$. Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is a basis for V. Let

$$\mathbf{v}_1 = \mathbf{x}_1,$$

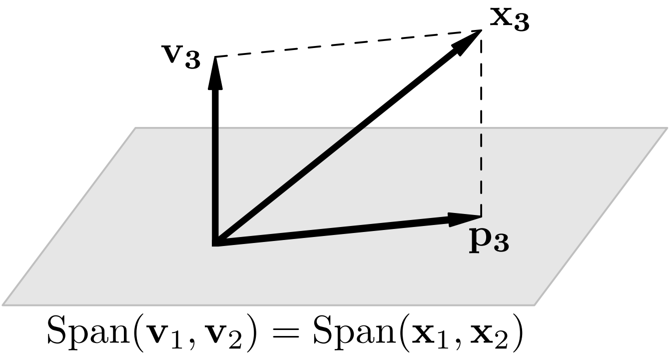
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1,$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2,$$

. . . .

$$\mathbf{v}_k = \mathbf{x}_k - \frac{\langle \mathbf{x}_k, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \cdots - \frac{\langle \mathbf{x}_k, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle} \mathbf{v}_{k-1}.$$

Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal basis for V.


x3
v3
p3
Span(v1, v2) = Span(x1, x2)

Any basis		Orthogonal basis
$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$	$\longrightarrow$	$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$

Properties of the Gram-Schmidt process:

- $\mathbf{v}_j = \mathbf{x}_j - (\alpha_1 \mathbf{x}_1 + \cdots + \alpha_{j-1} \mathbf{x}_{j-1})$, $1 \leq j \leq k$;
- the span of $\mathbf{v}_1, \dots, \mathbf{v}_j$ is the same as the span of $\mathbf{x}_1, \dots, \mathbf{x}_j$;
- $\mathbf{v}_j$ is orthogonal to $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$;
- $\mathbf{v}_j = \mathbf{x}_j - \mathbf{p}_j$, where $\mathbf{p}_j$ is the orthogonal projection of the vector $\mathbf{x}_j$ on the subspace spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$;
- $\|\mathbf{v}_j\|$ is the distance from $\mathbf{x}_j$ to the subspace spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{j-1}$.

Normalization

Let V be a subspace of $\mathbb{R}^n$. Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal basis for V.

Let $\mathbf{w}_1 = \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$, $\mathbf{w}_2 = \dfrac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$,..., $\mathbf{w}_k = \dfrac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$.

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ is an orthonormal basis for V.

Theorem Any non-trivial subspace of $\mathbb{R}^n$ admits an orthonormal basis.

Orthogonalization / Normalization

An alternative form of the Gram-Schmidt process combines orthogonalization with normalization.

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is a basis for a subspace $V \subset \mathbb{R}^n$. Let

$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|},$

$\mathbf{v}_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{w}_1 \rangle \mathbf{w}_1, \quad \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|},$

$\mathbf{v}_3 = \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{x}_3, \mathbf{w}_2 \rangle \mathbf{w}_2, \quad \mathbf{w}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|},$

....

$\mathbf{v}_k = \mathbf{x}_k - \langle \mathbf{x}_k, \mathbf{w}_1 \rangle \mathbf{w}_1 - \cdots - \langle \mathbf{x}_k, \mathbf{w}_{k-1} \rangle \mathbf{w}_{k-1},$
$\mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}.$

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ is an orthonormal basis for V.

Problem. Let Π be the plane spanned by vectors $\mathbf{x}_1 = (1,1,0)$ and $\mathbf{x}_2 = (0,1,1)$.

(i) Find the orthogonal projection of the vector $\mathbf{y} = (4,0,-1)$ onto the plane Π.

(ii) Find the distance from $\mathbf{y}$ to Π.

First we apply the Gram-Schmidt process to the basis $\mathbf{x}_1, \mathbf{x}_2$:

$$\mathbf{v}_1 = \mathbf{x}_1 = (1,1,0),$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (0,1,1) - \frac{1}{2}(1,1,0) = (-1/2, 1/2, 1).$$

Now that $\mathbf{v}_1, \mathbf{v}_2$ is an orthogonal basis for Π, the orthogonal projection of $\mathbf{y}$ onto Π is

$$\mathbf{p} = \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 = \frac{4}{2}(1,1,0) + \frac{-3}{3/2}(-1/2, 1/2, 1)$$

$$= (2,2,0) + (1,-1,-2) = (3,1,-2).$$

The distance from $\mathbf{y}$ to Π is $\|\mathbf{y} - \mathbf{p}\| = \|(1,-1,1)\| = \sqrt{3}$.