MATH 323

Linear Algebra

Lecture 21:
The Gram-Schmidt orthogonalization process.

Orthogonal sets

Let V be an inner product space with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

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

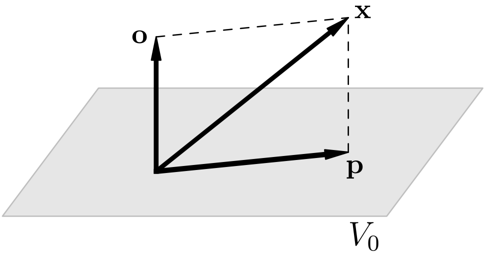
If, in addition, all vectors are of unit norm, $\|\mathbf{v}_i\| = 1$, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is called an **orthonormal set**.

Theorem Any orthogonal set is linearly independent.

Orthogonal projection

Theorem Let V be an inner product space and V_0 be a finite-dimensional subspace of V. Then any vector $\mathbf{x} \in V$ is uniquely represented as $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in V_0$ and $\mathbf{o} \perp V_0$.

The component $\mathbf{p}$ is called the **orthogonal projection** of the vector $\mathbf{x}$ onto the subspace V_0.



The projection $\mathbf{p}$ is closer to $\mathbf{x}$ than any other vector in V_0. Hence the distance from $\mathbf{x}$ to V_0 is $\|\mathbf{x} - \mathbf{p}\| = \|\mathbf{o}\|$.

Let V be an inner product space. Let $\mathbf{p}$ be the orthogonal projection of a vector $\mathbf{x} \in V$ onto a finite-dimensional subspace V_0.

If V_0 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 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is an orthogonal basis for V_0 then

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

Indeed, $\langle \mathbf{p}, \mathbf{v}_i \rangle = \displaystyle\sum_{j=1}^{n} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \langle \mathbf{v}_j, \mathbf{v}_i \rangle = \frac{\langle \mathbf{x}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \langle \mathbf{x}, \mathbf{v}_i \rangle$

$\implies \langle \mathbf{x} - \mathbf{p}, \mathbf{v}_i \rangle = 0 \implies \mathbf{x} - \mathbf{p} \perp \mathbf{v}_i \implies \mathbf{x} - \mathbf{p} \perp V_0.$

The Gram-Schmidt orthogonalization process

Let V be a vector space with an inner product. Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for V. Let

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

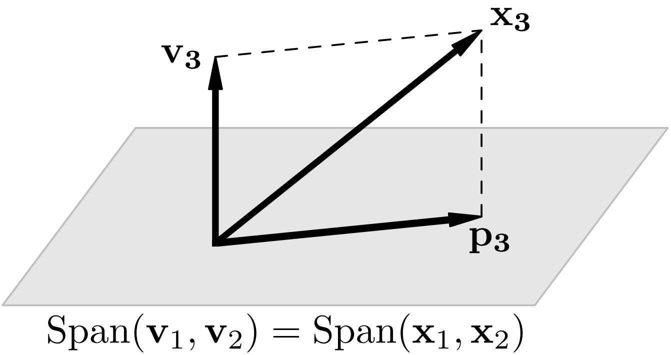
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1,$$

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

. . . .

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

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


$\mathbf{x_3}$
$\mathbf{v_3}$
$\mathbf{p_3}$
$\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2)$

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

Properties of the Gram-Schmidt process:

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

Normalization

Let V be a vector space with an inner product. Suppose $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is an orthogonal basis for V.

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

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

Theorem Any finite-dimensional vector space with an inner product has an orthonormal basis.

Remark. An infinite-dimensional vector space with an inner product may or may not have an orthonormal basis.

Orthogonalization / Normalization

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

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for an inner product space V. Let

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

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

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

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

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

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

Problem. Let V_0 be a subspace of dimension k in $\mathbb{R}^n$. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be a basis for V_0.

(i) Find an orthogonal basis for V_0.

(ii) Extend it to an orthogonal basis for $\mathbb{R}^n$.

Approach 1. Extend $\mathbf{x}_1, \ldots, \mathbf{x}_k$ to a basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ for $\mathbb{R}^n$. Then apply the Gram-Schmidt process to the extended basis. We shall obtain an orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ for $\mathbb{R}^n$. By construction, $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_k) = V_0$. It follows that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is a basis for V_0. Clearly, it is orthogonal.

Approach 2. First apply the Gram-Schmidt process to $\mathbf{x}_1, \ldots, \mathbf{x}_k$ and obtain an orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ for V_0. Secondly, find a basis $\mathbf{y}_1, \ldots, \mathbf{y}_m$ for the orthogonal complement $V_0^\perp$ and apply the Gram-Schmidt process to it obtaining an orthogonal basis $\mathbf{u}_1, \ldots, \mathbf{u}_m$ for $V_0^\perp$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_1, \ldots, \mathbf{u}_m$ is an orthogonal basis for $\mathbb{R}^n$.

Problem. Let Π be the plane in $\mathbb{R}^3$ spanned by vectors $\mathbf{x}_1 = (1, 2, 2)$ and $\mathbf{x}_2 = (-1, 0, 2)$.

(i) Find an orthonormal basis for Π.

(ii) Extend it to an orthonormal basis for $\mathbb{R}^3$.

$\mathbf{x}_1, \mathbf{x}_2$ is a basis for the plane Π. We can extend it to a basis for $\mathbb{R}^3$ by adding one vector from the standard basis. For instance, vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3 = (0, 0, 1)$ form a basis for $\mathbb{R}^3$ because

$$\begin{vmatrix} 1 & 2 & 2 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ -1 & 0 \end{vmatrix} = 2 \neq 0.$$

Using the Gram-Schmidt process, we orthogonalize the basis $\mathbf{x}_1 = (1, 2, 2)$, $\mathbf{x}_2 = (-1, 0, 2)$, $\mathbf{x}_3 = (0, 0, 1)$:

$$\mathbf{v}_1 = \mathbf{x}_1 = (1, 2, 2),$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (-1, 0, 2) - \frac{3}{9}(1, 2, 2)$$

$$= (-4/3, -2/3, 4/3),$$

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

$$= (0, 0, 1) - \frac{2}{9}(1, 2, 2) - \frac{4/3}{4}(-4/3, -2/3, 4/3)$$

$$= (2/9, -2/9, 1/9).$$

Now $\mathbf{v}_1 = (1,2,2)$, $\mathbf{v}_2 = (-4/3, -2/3, 4/3)$, $\mathbf{v}_3 = (2/9, -2/9, 1/9)$ is an orthogonal basis for $\mathbb{R}^3$ while $\mathbf{v}_1, \mathbf{v}_2$ is an orthogonal basis for Π. It remains to normalize these vectors.

$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = 9 \implies \|\mathbf{v}_1\| = 3$$
$$\langle \mathbf{v}_2, \mathbf{v}_2 \rangle = 4 \implies \|\mathbf{v}_2\| = 2$$
$$\langle \mathbf{v}_3, \mathbf{v}_3 \rangle = 1/9 \implies \|\mathbf{v}_3\| = 1/3$$

$$\mathbf{w}_1 = \mathbf{v}_1/\|\mathbf{v}_1\| = (1/3, 2/3, 2/3) = \tfrac{1}{3}(1,2,2),$$
$$\mathbf{w}_2 = \mathbf{v}_2/\|\mathbf{v}_2\| = (-2/3, -1/3, 2/3) = \tfrac{1}{3}(-2,-1,2),$$
$$\mathbf{w}_3 = \mathbf{v}_3/\|\mathbf{v}_3\| = (2/3, -2/3, 1/3) = \tfrac{1}{3}(2,-2,1).$$

$\mathbf{w}_1, \mathbf{w}_2$ is an orthonormal basis for Π.
$\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ is an orthonormal basis for $\mathbb{R}^3$.

Problem. Find the distance from the point $\mathbf{y} = (0, 0, 0, 1)$ to the subspace $V \subset \mathbb{R}^4$ spanned by vectors $\mathbf{x}_1 = (1, -1, 1, -1)$, $\mathbf{x}_2 = (1, 1, 3, -1)$, and $\mathbf{x}_3 = (-3, 7, 1, 3)$.

First we apply the Gram-Schmidt process to vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ to obtain an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ for the subspace V. Next we compute the orthogonal projection $\mathbf{p}$ of the vector $\mathbf{y}$ onto V:

$$\mathbf{p} = \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 + \frac{\langle \mathbf{y}, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_3, \mathbf{v}_3 \rangle} \mathbf{v}_3.$$

Then the distance from $\mathbf{y}$ to V equals $\|\mathbf{y} - \mathbf{p}\|$.

Alternatively, we can apply the Gram-Schmidt process to vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}$. We should obtain an orthogonal system $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$. By construction, $\mathbf{v}_4 = \mathbf{y} - \mathbf{p}$ so that the desired distance will be $\|\mathbf{v}_4\|$.

$\mathbf{x}_1 = (1, -1, 1, -1),\ \mathbf{x}_2 = (1, 1, 3, -1),$
$\mathbf{x}_3 = (-3, 7, 1, 3),\ \mathbf{y} = (0, 0, 0, 1).$

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

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (1, 1, 3, -1) - \frac{4}{4}(1, -1, 1, -1)$$

$$= (0, 2, 2, 0),$$

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

$$= (-3, 7, 1, 3) - \frac{-12}{4}(1, -1, 1, -1) - \frac{16}{8}(0, 2, 2, 0)$$

$$= (0, 0, 0, 0).$$

The Gram-Schmidt process can be used to check linear independence of vectors! It failed because the vector $\mathbf{x}_3$ is a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. V is a plane, not a 3-dimensional subspace. To fix things, it is enough to drop $\mathbf{x}_3$, i.e., we should orthogonalize vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}$.

$$\tilde{\mathbf{v}}_3 = \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2$$
$$= (0,0,0,1) - \frac{-1}{4}(1,-1,1,-1) - \frac{0}{8}(0,2,2,0)$$
$$= (1/4, -1/4, 1/4, 3/4).$$

$$|\tilde{\mathbf{v}}_3| = \left|\left(\frac{1}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}\right)\right| = \frac{1}{4}\,|(1,-1,1,3)| = \frac{\sqrt{12}}{4} = \frac{\sqrt{3}}{2}.$$

Problem. Find the distance from the point $\mathbf{z} = (0, 0, 1, 0)$ to the plane Π that passes through the point $\mathbf{x}_0 = (1, 0, 0, 0)$ and is parallel to the vectors $\mathbf{v}_1 = (1, -1, 1, -1)$ and $\mathbf{v}_2 = (0, 2, 2, 0)$.

The plane Π is not a subspace of $\mathbb{R}^4$ as it does not pass through the origin. Let $\Pi_0 = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. Then $\Pi = \Pi_0 + \mathbf{x}_0$.

Hence the distance from the point $\mathbf{z}$ to the plane Π is the same as the distance from the point $\mathbf{z} - \mathbf{x}_0$ to the plane $\Pi - \mathbf{x}_0 = \Pi_0$.

We shall apply the Gram-Schmidt process to vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{z} - \mathbf{x}_0$. This will yield an orthogonal system $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. The desired distance will be $\|\mathbf{w}_3\|$.

$\mathbf{v}_1 = (1, -1, 1, -1)$, $\mathbf{v}_2 = (0, 2, 2, 0)$, $\mathbf{z} - \mathbf{x}_0 = (-1, 0, 1, 0)$.

$$\mathbf{w}_1 = \mathbf{v}_1 = (1, -1, 1, -1),$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 = \mathbf{v}_2 = (0, 2, 2, 0) \ \text{ as } \mathbf{v}_2 \perp \mathbf{v}_1.$$

$$\begin{aligned}\mathbf{w}_3 &= (\mathbf{z} - \mathbf{x}_0) - \frac{\langle \mathbf{z} - \mathbf{x}_0, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{z} - \mathbf{x}_0, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \\ &= (-1, 0, 1, 0) - \frac{0}{4}(1, -1, 1, -1) - \frac{2}{8}(0, 2, 2, 0) \\ &= (-1, -1/2, 1/2, 0).\end{aligned}$$

$$|\mathbf{w}_3| = \left|\left(-1, -\frac{1}{2}, \frac{1}{2}, 0\right)\right| = \frac{1}{2}\,|(-2, -1, 1, 0)| = \frac{\sqrt{6}}{2} = \sqrt{\frac{3}{2}}.$$

Problem. Approximate the function $f(x) = e^x$ on the interval $[-1, 1]$ by a quadratic polynomial.

The best approximation would be a polynomial $p(x)$ that minimizes the distance relative to the uniform norm:

$$\|f - p\|_\infty = \max_{|x|\leq 1} |f(x) - p(x)|.$$

However there is no analytic way to find such a polynomial. Instead, one can find a *"least squares"* approximation that minimizes the integral norm

$$\|f - p\|_2 = \left(\int_{-1}^{1} |f(x) - p(x)|^2 \, dx \right)^{1/2}.$$

The norm $\|\cdot\|_2$ is induced by the inner product

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

Therefore $\|f - p\|_2$ is minimal if p is the orthogonal projection of the function f on the subspace $\mathcal{P}_3$ of polynomials of degree at most 2.

We should apply the Gram-Schmidt process to the polynomials $1, x, x^2,$ which form a basis for $\mathcal{P}_3$. This would yield an orthogonal basis p_0, p_1, p_2. Then

$$p(x) = \frac{\langle f, p_0\rangle}{\langle p_0, p_0\rangle}p_0(x) + \frac{\langle f, p_1\rangle}{\langle p_1, p_1\rangle}p_1(x) + \frac{\langle f, p_2\rangle}{\langle p_2, p_2\rangle}p_2(x).$$