Math 571
Winter 2004

Orthogonal Projections and Least Squares

1. Preliminaries

We start out with some background facts involving subspaces and inner products.

Definition 1.1. *Let U and V be subspaces of a vector space W such that $U \cap V = \{\mathbf{0}\}$. The* **direct sum** *of U and V is the set $U \oplus V = \{\mathbf{u} + \mathbf{v} \mid \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$.*

Definition 1.2. *Let S be a subspace of the inner product space V. The the* **orthogonal complement** *of S is the set $S^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{s} \rangle = 0 \text{ for all } \mathbf{s} \in S\}$.*

Theorem 1.3.
(1) *If U and V are subspaces of a vector space W with $U \cap V = \{\mathbf{0}\}$, then $U \oplus V$ is also a subspace of W.*
(2) *If S is a subspace of the inner product space V, then $S^\perp$ is also a subspace of V.*

Proof: (1.) Note that $\mathbf{0}+\mathbf{0} = \mathbf{0}$ is in $U \oplus V$. Now suppose $\mathbf{w}_1, \mathbf{w}_2 \in U \oplus V$, then $\mathbf{w}_1 = \mathbf{u}_1 + \mathbf{v}_1$ and $\mathbf{w}_2 = \mathbf{u}_2 + \mathbf{v}_2$ with $\mathbf{u}_i \in U$ and $\mathbf{v}_i \in V$ and $\mathbf{w}_1 + \mathbf{w}_2 = (\mathbf{u}_1 + \mathbf{v}_1) + (\mathbf{u}_2 + \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{v}_1 + \mathbf{v}_2)$. Since U and V are subspaces, it follows that $\mathbf{w}_1 + \mathbf{w}_2 \in U \oplus V$. Suppose now that α is a scalar, then $\alpha\mathbf{w}_1 = \alpha(\mathbf{u}_1 + \mathbf{v}_1) = \alpha\mathbf{u}_1 + \alpha\mathbf{v}_1$. As above, it then follows that $\alpha\mathbf{w}_1 \in U \oplus V$. Thus $U \oplus V$ is a subspace for W.

For (2.), first note that $\mathbf{0} \in S^\perp$. Now suppose that $\mathbf{v}_1$ and $\mathbf{v}_2 \in S^\perp$. Then $\langle \mathbf{v}_1, \mathbf{s} \rangle = \langle \mathbf{v}_2, \mathbf{s} \rangle = 0$ for all $\mathbf{s} \in S$. So $\langle \mathbf{v}_1 + \mathbf{v}_2, s \rangle = \langle \mathbf{v}_1, \mathbf{s} \rangle + \langle \mathbf{v}_2, \mathbf{s} \rangle = 0 + 0 = 0$ for all $\mathbf{s} \in S$. Thus $\mathbf{v}_1 + \mathbf{v}_2 \in S^\perp$. Similarly, if α is a scalar, then $\langle \alpha\mathbf{v}_1, s \rangle = \alpha\langle \mathbf{v}_1, s \rangle = \alpha \cdot 0 = 0$ for all $\mathbf{s} \in S$. Thus $S^\perp$ is a subspace of V. □

Theorem 1.4. *If U and V are subspaces of W with $U \cap V = \{\mathbf{0}\}$ and $\mathbf{w} \in U \oplus V$, then $\mathbf{w} = \mathbf{u} + \mathbf{v}$ for unique $\mathbf{u} \in U$ and $\mathbf{v} \in V$.*

Proof: Write $\mathbf{w} = \mathbf{u}_1 + \mathbf{v}_1$ and $\mathbf{w} = \mathbf{u}_2 + \mathbf{v}_2$. Then $\mathbf{u}_1 + \mathbf{v}_1 = \mathbf{u}_2 + \mathbf{v}_2 \Rightarrow \mathbf{u}_1 - \mathbf{u}_2 = \mathbf{v}_2 - \mathbf{v}_1 \Rightarrow \mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0} = \mathbf{v}_2 - \mathbf{v}_1 \Rightarrow \mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{v}_2 = \mathbf{v}_1$. □

Recall that one of the axioms of an inner product is that $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$. An immediate consequence of this is that $S \cap S^\perp = 0$.

Definition 1.5. *Let S be a subspace of the inner product space V and let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a basis for S such that $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ if $i \neq j$, then this basis is called an* **orthogonal basis**. *Furthermore, if $\langle \mathbf{x}_i, \mathbf{x}_i \rangle = 1$ then this basis is called an* **orthonormal basis**.

Definition 1.6. *Let S be a finite dimensional subspace of the inner product space V and let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an orthogonal basis for S. If $\mathbf{v}$ is any vector in V then the* **orthogonal projection** *of $\mathbf{v}$ onto S is the vector:*

$$\mathbf{p} = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle} \mathbf{x}_i$$

Note that if $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is an orthonormal basis, then we have the simpler expression:

$$\mathbf{p} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{x}_i \rangle \mathbf{x}_i$$

Also in the special case where S is spanned be the single vector $\mathbf{x}_1$, then $\mathbf{p}$ is just the usual orthogonal projection of $\mathbf{v}$ onto S, which is the line spanned by $\mathbf{x}_1$.

Now we can prove the main theorem of this section:

Theorem 1.7. *Let S be a finite dimensional subspace of the inner product space V and $\mathbf{v}$ be some vector in V. Moreover let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be an orthogonal basis for S and $\mathbf{p}$ be the orthogonal projection of $\mathbf{v}$ onto S. Then*

(1) $\mathbf{v} - \mathbf{p} \in S^{\perp}$.
(2) $V = S \oplus S^{\perp}$.
(3) *If $\mathbf{y}$ is any vector in S with $\mathbf{y} \neq \mathbf{p}$, then* $||\mathbf{v} - \mathbf{p}|| < ||\mathbf{v} - \mathbf{y}||$

Note that part (3.) says that $\mathbf{p}$ is the vector in S which is closest to $\mathbf{v}$. Moreover, an immediate consequence of (2.) is that the orthogonal projection $\mathbf{p}$ of $\mathbf{v}$ onto S is independent of the choice of orthogonal basis for S.

Proof: (1.) We need to show that $\mathbf{p}$ and $\mathbf{v} - \mathbf{p}$ are orthogonal. So consider $\langle \mathbf{p}, \mathbf{v} - \mathbf{p} \rangle = \langle \mathbf{p}, \mathbf{v} \rangle - \langle \mathbf{p}, \mathbf{p} \rangle$. Note that $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ when $i \neq j$, so that

$$\langle \mathbf{p}, \mathbf{v} \rangle = \sum_{i=1}^{n} \langle c_i \mathbf{x}_i, \mathbf{v} \rangle \text{ with } c_i = \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle} \text{ and } \langle \mathbf{p}, \mathbf{p} \rangle = \sum_{i=1}^{n} \langle c_i \mathbf{x}_i, c_i \mathbf{x}_i \rangle \Rightarrow$$

$$\langle \mathbf{p}, \mathbf{v} \rangle = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle} \langle \mathbf{x}_i, \mathbf{v} \rangle \text{ and } \langle \mathbf{p}, \mathbf{p} \rangle = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle^2}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle^2} \langle \mathbf{x}_i, \mathbf{x}_i \rangle$$

Thus

$$\langle \mathbf{p}, \mathbf{v} \rangle = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle^2}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle} \text{ and } \langle \mathbf{p}, \mathbf{p} \rangle = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{x}_i \rangle^2}{\langle \mathbf{x}_i, \mathbf{x}_i \rangle}$$

and the result follows for this part. Now let $\mathbf{v}$ be any vector in V, then $\mathbf{v} = \mathbf{p} + (\mathbf{v} - \mathbf{p})$. Note that $\mathbf{p} \in S$ and from (1.), $\mathbf{v} - \mathbf{p} \in S^{\perp}$, and $S \cap S^{\perp} = \{\mathbf{0}\}$. Therefore we must have $V = S \oplus S^{\perp}$, proving (2.). For part (3.), let $\mathbf{y}$ be some vector in S with $\mathbf{y} \neq \mathbf{p}$. Then $||\mathbf{v} - \mathbf{p}|| = ||\mathbf{v} - \mathbf{p} + \mathbf{p} - \mathbf{y}||$. Since $\mathbf{p} - \mathbf{y} \in S$ and $\mathbf{v} - \mathbf{p} \in S^{\perp}$ by (1.), we have

$$||\mathbf{v} - \mathbf{p}||^2 + ||\mathbf{p} - \mathbf{y}||^2 = ||\mathbf{v} - \mathbf{y}||^2$$

By the Pythagorean Theorem. So

$$||\mathbf{v} - \mathbf{p}||^2 = ||\mathbf{v} - \mathbf{y}||^2 - ||\mathbf{p} - \mathbf{y}||^2$$

Since $\mathbf{y} \neq \mathbf{p}$, $||\mathbf{p} - \mathbf{y}|| \neq 0$. Therefore $||\mathbf{v} - \mathbf{p}||^2 < ||\mathbf{v} - \mathbf{y}||^2$ and $||\mathbf{v} - \mathbf{y}|| < ||\mathbf{v} - \mathbf{y}||$. □

Note that by (3.) of the above theorem, if $\mathbf{v}$ is actually in S, then $\mathbf{p} = \mathbf{v}$.

Definition 1.8. *Let S be a subspace of the inner product space V, $\mathbf{v}$ be a vector in V and $\mathbf{p}$ be the orthogonal projection of $\mathbf{v}$ onto S. Then $\mathbf{p}$ is called the* **least squares approximation** *of $\mathbf{v}$ (in S) and the vector $\mathbf{r} = \mathbf{v} - \mathbf{p}$ is called the* **residual vector** *of $\mathbf{v}$.*

2. Least squares in $\mathbb{R}^n$

In this section we consider the following situation: Suppose that A is an $m \times n$ real matrix with $m > n$. If $\mathbf{b}$ is a vector in $\mathbb{R}^m$ then the matrix equation $A\mathbf{x} = \mathbf{b}$ corresponds to an overdetermined linear system. Generally such a system does not have a solution, however we would like to find an $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is as close to $\mathbf{b}$ as possible. In this case $A\hat{\mathbf{x}}$ is the least squares approximation to $\mathbf{b}$ and we refer to $\hat{\mathbf{x}}$ as the **least squares solution** to this system. Recall that if $\mathbf{r} = \mathbf{b} - A\mathbf{x}$, then $\mathbf{r}$ is the residual of this system. Moreover, our goal is then to find a $\mathbf{x}$ which minimizes $||\mathbf{r}||$.

Before we continue, we mention a result without proof:

Theorem 2.1. *Suppose that A is a real matrix. Then* $\mathrm{Col}(A)^{\perp} = \mathrm{N}(A^T)$ *and* $\mathrm{Col}(A^T)^{\perp} = \mathrm{N}(A)$.

We will use the results of the previous section to find $\hat{\mathbf{x}}$, or more precisely $A\hat{\mathbf{x}}$. Given $\mathbf{b}$ there is a unique vector $\mathbf{p}$ in $\mathrm{Col}(A)$ such that $||\mathbf{b} - \mathbf{p}||$ is minimal by theorem 1.7. Moreover, by the same theorem, $\mathbf{b} - \mathbf{p} \in \mathrm{N}(A^T)$. Thus:

$$A^T(\mathbf{b} - \mathbf{p}) = \mathbf{0} \Rightarrow A^T\mathbf{b} - A^T\mathbf{p} = \mathbf{0} \Rightarrow A^T\mathbf{p} = A^T\mathbf{b}$$

However, $\mathbf{p} = A\hat{\mathbf{x}}$ for some vector $\hat{\mathbf{x}}$ (note: $\hat{\mathbf{x}}$ is not necessarily unique, but $A\hat{\mathbf{x}}$ is). So

$$A^T\mathbf{p} = A^T\mathbf{b} \Rightarrow A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

Thus to find $\hat{\mathbf{x}}$ we simply solve for $\hat{\mathbf{x}}$ in the equation:

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

which is necessarily consistent. Note that in this case we did not need to know an orthogonal basis for $\mathrm{Col}(A)$. This is because we never explicitly calculate $\mathbf{p}$.

Another general fact about A in this case is that the rank of A is generally n. That is, the columns of A will usually be linearly independent. We have the following theorem which gives us an additional way to solve for $\hat{\mathbf{x}}$ in this situation:

Theorem 2.2. *If A is an $m \times n$ matrix and the rank of A is n then A^TA is invertible.*

Proof: Clearly, $\mathrm{N}(A)$ is a subset of $\mathrm{N}(A^TA)$. We now wish to show that $\mathrm{N}(A^TA)$ is a subset of $\mathrm{N}(A)$. This would establish that $\mathrm{N}(A) = \mathrm{N}(A^TA)$. Let $\mathbf{x} \in \mathrm{N}(A^TA)$, then $(A^TA)\mathbf{x} = 0 \Rightarrow A^T(A\mathbf{x}) = 0 \Rightarrow A\mathbf{x} \in \mathrm{N}(A^T)$. Note also that $A\mathbf{x} \in \mathrm{Col}(A)$ so that $A\mathbf{x} \in \mathrm{N}(A^T) \cap \mathrm{Col}(A)$. Since $\mathrm{Col}(A)^{\perp} = \mathrm{N}(A^T) \Rightarrow A\mathbf{x} = \mathbf{0}$, thus $x \in \mathrm{N}(A)$ and $\mathrm{N}(A^TA) = \mathrm{N}(A)$. By the rank-nullity theorem we see that the rank of A^TA is the same as the rank of A which is assumed to be n. As A^TA is an $n \times n$ matrix, it must be invertible. □

Thus, when A has rank n, A^TA is invertible, and

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$$

Now we proceed with some examples:

Example 1: Consider the linear system:

$$\begin{aligned} -x_1 + x_2 &= 10 \\ 2x_1 + x_2 &= 5 \\ x_1 - 2x_2 &= 20 \end{aligned}$$

This system is overdetermined and inconsistent. We would like to find the least squares approximation to $\mathbf{b}$ and the least squares solution $\hat{\mathbf{x}}$ to this system. We can rewrite this linear system as a matrix system $A\mathbf{x} = \mathbf{b}$ where:

$$A = \begin{pmatrix} -1 & 1 \\ 2 & 1 \\ 1 & -2 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 10 \\ 5 \\ 20 \end{pmatrix}$$

It is easy to check that A has rank 2, hence A^TA is invertible. Therefore:

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \begin{pmatrix} 2.71 \\ -3.71 \end{pmatrix}, \qquad A\hat{\mathbf{x}} = \begin{pmatrix} -6.43 \\ 1.71 \\ 10.14 \end{pmatrix} \text{ and } ||\mathbf{r}|| = ||\mathbf{b} - A\hat{\mathbf{x}}|| = 19.44$$

Example 2: Suppose some system is modeled by a quadratic function $f(x)$, so that $f(x) = ax^2 + bx + c$. Experimental data is recorded in the form $(x, f(x))$ with the following results:

$$(1,1), (2,10), (3,9), (4,16)$$

We would like to find the best approximation for $f(x)$. Using these data points, we see that:

$$\begin{aligned} a(1) + b(1) + c &= 1 \\ a(4) + b(2) + c &= 10 \\ a(9) + b(3) + c &= 9 \\ a(16) + b(4) + c &= 16 \end{aligned}$$

This corresponds to the matrix equation $A\mathbf{x} = \mathbf{b}$ where:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 1 \\ 10 \\ 9 \\ 16 \end{pmatrix}$$

As in the previous example, the matrix A has full rank, hence $A^T A$ is invertible. Therefore the least squares solution to this system is:

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = \begin{pmatrix} -0.5 \\ 6.9 \\ -4.5 \end{pmatrix}$$

Therefore $f(x)$ is approximately $-0.5x^2 + 6.9x - 4.5$

Example 3: The orbit of a comet around the sun is either elliptical, parabolic, or hyperbolic. In particular, the orbit can be expressed by the polar equation:

$$r = \beta - e(r\cos\theta)$$

where β is some positive constant and e is the eccentricity. Note that the orbit is elliptical if $0 \leq e < 1$, parabolic if $e = 1$ and hyperbolic if $e > 1$.

A certain comet is observed over time and the following set of data (r, θ) was recorded:

$$(1.2, 0.3), (2.1, 1.2), (4.1, 2.6), (6.3, 3.8)$$

Using this data we would like to find the approximate orbital equation of this comet. Plugging these data points in the equation above gives us the linear system:

$$\begin{aligned} \beta - e(1.146) &= 1.2 \\ \beta - e(0.761) &= 2.1 \\ \beta - e(-3.513) &= 4.1 \\ \beta - e(-4.983) &= 6.3 \end{aligned}$$

This system corresponds to the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 1 & -1.146 \\ 1 & -0.761 \\ 1 & 3.513 \\ 1 & 4.983 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 1.2 \\ 2.1 \\ 4.1 \\ 6.3 \end{pmatrix}$$

Once again, the matrix A has full rank so that the least squares solution is given by:

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = \begin{pmatrix} 2.242 \\ 0.718 \end{pmatrix}$$

Therefore the orbital equation is approximately $r = 2.242 - 0.718(r\cos\theta)$. This example is similar to one of the first applications of least squares. Gauss is credited with developing the method of least squares and applying it to predicting the path of the asteroid Ceres. He did this to a remarkable degree of accuracy in approximately 1801. In particular, he predicted where the asteroid would be in 40 days after it had passed behind the sun.

3. Least squares in $C[a, b]$

Recall that an inner product in $C[a, b]$ is given by

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

where $w(x)$ is some continuous, positive function on $[a, b]$. Consider that we have a collection of functions $\{f_1(x), \ldots, f_n(x)\}$ which are mutually orthogonal. Moreover, assume that they form an orthogonal basis for S. Then, given any function $f(x)$ in $C[a, b]$, we can approximate $f(x)$ by a linear combination of the f_i. The best such approximation (in terms of least squares) will be given by the orthogonal projection $\mathbf{p}(x)$ of $f(x)$ onto S. The most common application of such an approximation is in Fourier Series which will be covered in the next section.

There is an analytical consideration which has to be made, and that is how good can we make this approximation. In particular, can we enlarge S is a regular way so that the limit of this process is $f(x)$. This is a question which is beyond our scope, but the answers is yes in some cases and no in others.

4. Fourier Series

In this section we consider the function space $C[-\pi, \pi]$ and we wish to know if given a function $f(x)$ in $C[-\pi, \pi]$ how can we approximate this function using functions of the form $\cos mx$ and $\sin mx$. This is obviously useful for periodic functions. Our setup is as follows:

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

and

$$S_n = \text{Span}(\frac{1}{\sqrt{2}}, \sin x, \ldots, \sin nx, \cos x, \ldots, \cos nx)$$

We can check that the following are true:

$$\langle 1, \sin mx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin mx\ dx = 0$$

$$\langle 1, \cos mx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos mx\ dx = 0$$

$$\langle \cos mx, \cos kx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos mx \cos kx\ dx = \begin{cases} 1 \text{ if } m = k \\ 0 \text{ if } m \neq k \end{cases}$$

$$\langle \sin mx, \sin kx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin mx \sin kx\ dx = \begin{cases} 1 \text{ if } m = k \\ 0 \text{ if } m \neq k \end{cases}$$

$$\langle \cos mx, \sin kx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos mx \sin kx\ dx = 0$$

Therefore

$$\{\frac{1}{\sqrt{2}}, \sin x, \dots, \sin nx, \cos x, \dots, \cos nx\}$$

is an orthonormal basis for S_n.

Given a function $f(x)$ in $C[-\pi, \pi]$, the least squares approximation of $f(x)$ in S_n will be

$$\frac{a_0}{2} + \sum_{i=1}^{n} (a_i \cos ix + b_i \sin ix)$$

where

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\ dx$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx\ dx$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx\ dx$$

Note that a_k and b_k are just the inner products of $f(x)$ with the basis vectors of S_n and

$$\frac{a_0}{2} + \sum_{i=1}^{n} (a_i \cos ix + b_i \sin ix)$$

is just the orthogonal projection of $f(x)$ onto S_n.

The series $\frac{a_0}{2} + \sum_{i=1}^{n} (a_i \cos ix + b_i \sin ix)$ is called the nth order or nth degree **Fourier series** approximation of $f(x)$ and a_i, b_i are called the **Fourier coefficients** . If we consider the approximations as partial sums, then as $n \to \infty$ we get the usual Fourier series $\frac{a_0}{2} + \sum_{i=1}^{\infty} (a_i \cos ix + b_i \sin ix)$. There still remains the question of how good an approximation this is and if this series actually converges to $f(x)$.

Now lets try see what the projection of an easy function onto S_n is:

Example 4: Let $f(x) = x$. Then clearly $f(x)$ is a vector in $C[-\pi, \pi]$. The Fourier coefficients are:

$$a_0 = \frac{1}{\pi} \int_{-pi}^{pi} x\ dx = 0$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} x \cos kx\ dx = \left. \frac{x \sin kx}{k\pi} \right|_{-\pi}^{\pi} - \frac{1}{k\pi} \int_{-\pi}^{\pi} \sin kx\ dx = 0$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} x \cos kx\ dx = \left. \frac{-x \cos kx}{k\pi} \right|_{-\pi}^{\pi} - \frac{1}{k\pi} \int_{-\pi}^{\pi} - \cos kx\ dx = (-1)^{k+1} \frac{2}{k}$$

Therfore the closest vector in S_n to $f(x)$ is

$$\mathbf{p}_n(x) = \sum_{i=1}^{n} b_k \sin ix = \sum_{i=1}^{n} (-1)^{i+1} \frac{2}{i} \sin ix$$

It is beyond our scope, but as $n \to \infty$, these approximations do converge to $f(x)$ on the interval $(-\pi, pi)$. In particular, on that interval,

$$x = f(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{2}{n} \sin nx$$

and

$$\frac{x}{2} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} \sin nx$$

If we take $x = \pi/2$, then we see that

$$\frac{\pi}{4} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} \sin n\pi = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$