Lecture no 8: APPLICATION OF ORTHOGONALITY: BEST APPROXIMATION; LEAST SQUARES

There are many applications in which some linear system $A\,x = b$ of m equations in n unknowns should be consistent on physical grounds but fails to be so because of *measurement errors* in the entries of A or b. In such cases one looks for vectors that come as close as possible to being solutions in the sense that they *minimize* $||b - A\,x||$ with respect to the Euclidean inner product on $\mathbf{R}^m$. In this section we will discuss methods for finding such minimizing vectors.

1. Best Approximation

Suppose that $A\,x = b$ is an inconsistent linear system[1] of m equations in n unknowns in which we suspect the inconsistency to be caused by errors in the entries of A or b. Since no exact solution is possible, we will look for a vector x that comes as "close as possible" to being a solution in the sense that it minimizes $||b - A\,x||$ with respect to the Euclidean inner product on $\mathbf{R}^m$.

You can think of $A\,x$ as an *approximation* to b and $||b - A\,x||$ as the *error* in that approximation—the smaller the error, the better the approximation. This leads to the following problem:

Least Squares Problem: Given a linear system $A\,x = b$ of m equations in n unknowns, find a vector x in $\mathbf{R}^n$ that minimizes $||b - A\,x||$ with respect to the Euclidean inner product on $\mathbf{R}^m$. We call such a vector, if it exists, a *least squares solution* of $A\,x = b$, we call $b - A\,x$ the *least squares error vector*, and we call $||b - A\,x||$ the *least squares error*.

To explain the terminology in this problem, suppose that the column form of $b - A\,x$ is

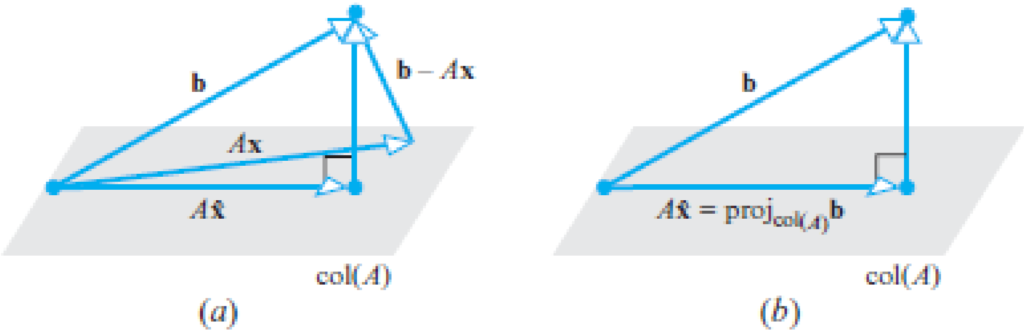
$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

The term "*least squares solution*" results from the fact that minimizing $b - A\,x$ also has the effect of minimizing $||b - A\,x||^2 = e_1^2 + e_2^2 + \cdots + e_m^2$

[1]***Remark:*** If a linear system is consistent, then its exact solutions are the same as its least squares solutions, in which case the least squares error is zero.

What is important to keep in mind about the *least squares problem* is that for every vector x in $\mathbf{R}^n$, the product $A\,x$ is in the column space of A because it is a linear combination of the column vectors of A. That being the case, to find a *least squares solution* of $A\,x = b$ is equivalent to finding a vector $A\hat{x}$ in the column space of A that is closest to b in the sense that it minimizes the length of the vector $b - A\,x$. This is illustrated in Figure, which also suggests that $A\hat{x}$ is the orthogonal projection of b on the column space of A, that is (Figure),

$$A\hat{x} = \mathbf{proj}(b, \text{col}(A))$$



The next theorem will confirm this conjecture.

Theorem: (**Best Approximation Theorem**). If W is a finite-dimensional subspace of an inner product space V, and if b is a vector in V, then $\mathbf{proj}(b, W)$ is the best approximation to b from W in the sense that

$$\|b - \mathbf{proj}(b, W)\| < \|b - w\|,$$

for every vector w in W that is different from $\mathbf{proj}(b, W)$.

Finding Least Squares Solutions

One way to find a least squares solution of $A\,x = b$ is to calculate the orthogonal projection $\mathbf{proj}(b, W)$ on the column space W of A and then solve the equation

$$A\,x = \mathbf{proj}(b, W)$$

However, we can avoid calculating the projection by rewriting the expression as

$$b - A\,x = b - \mathbf{proj}(b, W)$$

and then multiplying both sides of this equation by A^T to obtain

$$A^T(b - A\,x) = A^T(b - \mathbf{proj}(b, W))$$

Since $b - \mathbf{proj}(b, W)$ is the component of b that is orthogonal to the column space of A, it follows from Theorem that this vector lies in the null space of A^T, and hence that

$$A^T(b - \mathbf{proj}(b, W)) = 0$$

Thus, the expression simplifies to

$$A^T (b - A x) = 0$$

which we can rewrite as

$$A^T A x = A^T b$$

This is called the *normal equation* or the *normal system* associated with $Ax = b$. When viewed as a linear system, the individual equations are called the *normal equations* associated with $Ax = b$.

In summary, we have established the following result.

Theorem: For every linear system $A x = b$, the associated normal system

$$A^T A x = A^T b$$

is consistent, and all solutions of the system are least squares solutions of $Ax = b$. Moreover, if W is the column space of A, and x is any least squares solution of $Ax = b$, then the *orthogonal projection* of b onto W is

$$\mathbf{proj}(b, W)\,) = Ax$$

Example 1: (Unique Least Squares Solution) Find the least squares solution, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned}$$

It will be convenient to express the system in the matrix form $Ax = b$, where

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

It follows that

$$A^T A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}, \quad A^T \mathbf{b} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

so the normal system $A^T A x = A^T b$ is

$$\begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

Solving this system yields a unique least squares solution, namely,

$$x_1 = \tfrac{17}{95}, \quad x_2 = \tfrac{143}{285}$$

The least squares error vector is

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} -\frac{92}{285} \\ \frac{439}{285} \\ \frac{95}{57} \end{bmatrix} = \begin{bmatrix} \frac{1232}{285} \\ -\frac{154}{285} \\ \frac{4}{3} \end{bmatrix}$$

and the least squares error is

$$\|b - A\,x\| \cong 4.556$$

The computations in the next example are a little tedious for hand computation, so in absence of a calculating utility you may want to just read through it for its ideas and logical flow.

Example 2: (Infinitely Many Least Squares Solutions) Find the least squares solutions, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &= 2 \\ x_1 - 4x_2 + 3x_3 &= -2 \\ x_1 + 10x_2 - 7x_3 &= 1 \end{aligned}$$

The matrix form of the system is $Ax = b$, where

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

It follows that

$$A^TA = \begin{bmatrix} 11 & 12 & -7 \\ 12 & 120 & -84 \\ -7 & -84 & 59 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 22 \\ -15 \end{bmatrix}$$

so the normal system $A^T A\, x = A^T b$ is

$$\left[\begin{array}{ccc|c} 11 & 12 & -7 & 5 \\ 12 & 120 & -84 & 22 \\ -7 & -84 & 59 & -15 \end{array}\right]$$

The reduced row echelon form of this matrix is

$$\left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{7} & \frac{2}{7} \\ 0 & 1 & -\frac{5}{7} & \frac{13}{84} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

from which it follows that there are infinitely many least squares solutions, and that they are given by the parametric equations

$$x_1 = \tfrac{2}{7} - \tfrac{1}{7}t$$
$$x_2 = \tfrac{13}{84} + \tfrac{5}{7}t$$
$$x_3 = t$$

As a check, let us verify that all least squares solutions produce the same least squares error vector and the same least squares error. To see that this is so, we first compute

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \begin{bmatrix} \frac{2}{7} - \frac{1}{7}t \\ \frac{13}{84} + \frac{5}{7}t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ -\frac{1}{3} \\ \frac{11}{6} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ -\frac{5}{3} \\ -\frac{5}{6} \end{bmatrix}$$

Since $b - A\,x$ does not depend on t, all least squares solutions produce the same error vector, namely

$$\|\mathbf{b} - A\mathbf{x}\| = \sqrt{\left(\tfrac{5}{6}\right)^2 + \left(-\tfrac{5}{3}\right)^2 + \left(-\tfrac{5}{6}\right)^2} = \tfrac{5}{6}\sqrt{6}$$

Conditions for Uniqueness of Least Squares Solutions

We know from Theorem that the system $A^T A\,x = A^T b$ of normal equations that is associated with the system $A\,x = b$ is consistent. Thus, every linear system $A\,x = b$ has either one least squares solution (as in Example 1) or infinitely many least squares solutions (as in Example 2). Since $A^T A$ is a square matrix, the former occurs if $A^T A$ is invertible and the latter if it is not. The next two theorems are concerned with this idea.

***Theorem* 1:** If A is an $m \times n$ matrix, then the following are equivalent.

(a) The column vectors of A are linearly independent.

(b) $A^T A$ is invertible.

The next theorem gives an explicit formula for the least squares solution of a linear system in which the coefficient matrix has linearly independent column vectors.

***Theorem* 2:** If A is an $m \times n$ matrix with linearly independent column vectors, then for every $m \times 1$ matrix b, the linear system $A\,x = b$ has a unique least squares solution. This solution is given by

$$x = (A^T A)^{-1} A^T b$$

Moreover, if W is the column space of A, then the orthogonal projection of b on W is

$$\mathbf{proj}(b, W) = A\,x = A\,(A^T A)^{-1} A^T b$$

Example 3: (A Formula Solution to Example 1) Use the Theorem to find the least squares solution of the linear system in Example 1.

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

$$= \frac{1}{285}\begin{bmatrix} 21 & 3 \\ 3 & 14 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix}$$

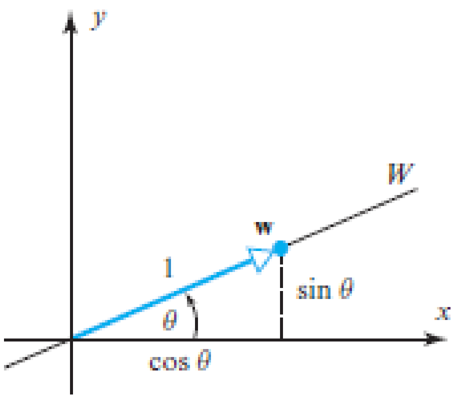
Note: The standard matrix for the *orthogonal projection* on the column space of a matrix A is

$$P = A\,(A^T A)^{-1}A^T$$

Example 4: (Orthogonal Projection on a Column Space) Earlier we showed that the standard matrix for the orthogonal projection onto the line W through the origin of $\mathbf{R}^2$ that makes an angle θ with the positive x-axis is

$$P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$$

Derive this result?



First we must find a matrix A for which the line W is the column space. Since the line is one-dimensional and consists of all scalar multiples of the vector w = (cos θ, sin θ) (see Figure), we can take A to be

$$A = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

Then,

$$A(A^TA)^{-1}A^T = AA^T = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} [\cos\theta \quad \sin\theta]$$

$$= \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} = P_\theta$$

2 Mathematical Modeling Using Least Squares

In this subsection we will use results about orthogonal projections in inner product spaces to obtain a technique for *fitting a line or other polynomial curve* to a set of experimentally determined points in the plane.

Fitting a Curve to Data

A common problem in experimental work is to obtain a mathematical relationship $y = f(x)$ between two variables x and y by "*fitting*" a curve to points in the plane corresponding to various experimentally determined values of x and y, say

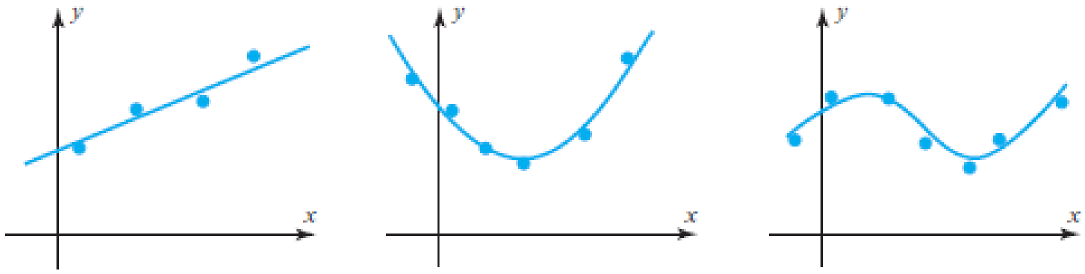
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

On the basis of theoretical considerations or simply by observing the pattern of the points, the experimenter decides on the general form of the curve $y = f(x)$ to be fitted. This curve is called a *mathematical model* of the data. Some examples are (Figure):

(a) A straight line: $y = a + bx$

(b) A quadratic polynomial: $y = a + bx + cx^2$

(c) A cubic polynomial: $y = a + bx + cx^2 + dx^3$



(a) $y = a + bx$ (b) $y = a + bx + cx^2$ (c) $y = a + bx + cx^2 + dx^3$

Least Squares Fit of a Straight Line

When data points are obtained experimentally, there is generally some measurement "*error*" making it impossible to find a curve of the desired form that passes through all the points. Thus, the idea is to choose the curve (by determining its coefficients) that "*best fits*" the data. We begin with the simplest case: fitting a straight line to data points.

Suppose we want to fit a straight line $y = a + bx$ to the experimentally determined points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

If the data points were collinear, the line would pass through all n points, and the unknown coefficients a and b would satisfy the equations

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned}$$

We can write this system in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or more compactly as $M\,v = y$, where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$$

If there are measurement errors in the data, then the data points will typically not lie on a line, and the system of linear equations will be inconsistent. In this case we look for a least squares approximation to the values of a and b by solving the normal system

$$M^{\mathrm{T}} M\, v = M^{\mathrm{T}}\, y,$$

For simplicity, let us assume that the x-coordinates of the data points are not all the same, so M has linearly independent column vectors and the normal system has the unique solution

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = (M^T M)^{-1} M^T \mathbf{y}$$

The line $y = a^* + b^* x$ that results from this solution is called the least squares line of best fit or the regression line. It follows that this line minimizes

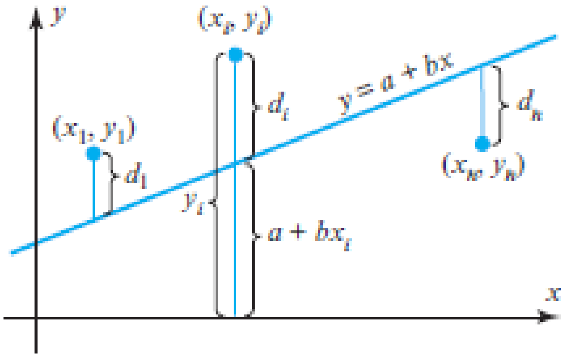
$$\|\mathbf{y} - M\mathbf{v}\|^2 = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2$$

The quantities

$$d_1 = |y_1 - (a + bx_1)|, \quad d_2 = |y_2 - (a + bx_2)|, \ldots, \quad d_n = |y_n - (a + bx_n)|$$

are called *residuals*.

Since the residual d_{i} is the distance between the data point (x_{i}, y_{i}) and the regression line (Figure), we can interpret its value as the “error” in y_{i} at the point x_{i}. If we assume that the value of each x_i is exact, then all the errors are in the y_{i} so the regression line can be described as the line that minimizes the sum of the squares of the data errors—hence the name, “*least squares line of best fit*”.



In summary, we have the following theorem.

***Theorem*:** (**Uniqueness of the Least Squares Solution**) Let (x_1, y_1), (x_2, y_2), . . . , (x_n, y_n) be a set of two or more data points, not all lying on a vertical line, and let

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \text{ and } y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then there is a *unique least squares* straight line fit $y = a^* + b^*x$ to the data points. Moreover,

$$v^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

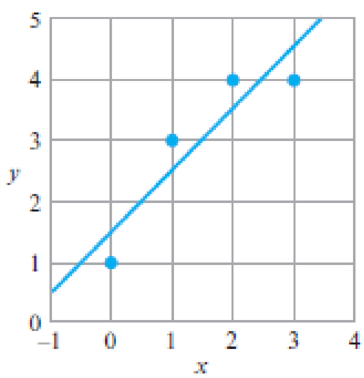
is given by the formula

$$v^* = (M^T M)^{-1} M^T y$$

which expresses the fact that $v = v^*$ is the unique solution of the normal equation

$$M^T M v = M^T y$$

Example 1: (Least Squares Straight Line Fit) Find the least squares straight line fit to the four points (0, 1), (1, 3), (2, 4), and (3, 4).



We have

$$M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad M^T M = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix},$$

and

$$(M^T M)^{-1} = \frac{1}{10} \begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}$$

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

so the desired line is $y = 1.5 + x$.

Example 2: (Fitting a Quadratic Curve to Data) According to Newton's second law of motion, a body near the Earth's surface falls vertically downward in accordance with the equation

$$s = s_0 + v_0t + \tfrac{1}{2}gt^2$$

where

s = vertical displacement downward relative to some reference point
s_0 = displacement from the reference point at time $t = 0$
v_0 = velocity at time $t = 0$
g = acceleration of gravity at the Earth's surface

Suppose that a laboratory experiment is performed to approximate g by measuring the displacement s relative to a fixed reference point of a falling weight at various times. Use the experimental results shown in the following table to approximate g.

Time t (sec)	.1	.2	.3	.4	.5
Displacement s (ft)	−0.18	0.31	1.03	2.48	3.73

For notational simplicity, let $a_0 = s_0$, $a_1 = v_0$, and $a_2 = 1/2g$ in the formula, so our mathematical problem is to fit a quadratic curve

$$s = a_0 + a_1t + a_2t^2$$

to the five data points:

$$(.1, -0.18), \quad (.2, 0.31), \quad (.3, 1.03), \quad (.4, 2.48), \quad (.5, 3.73)$$

With the appropriate adjustments in notation, the matrices M and $\mathbf{y}$ in (11) are

$$M = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} = \begin{bmatrix} 1 & .1 & .01 \\ 1 & .2 & .04 \\ 1 & .3 & .09 \\ 1 & .4 & .16 \\ 1 & .5 & .25 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{bmatrix} = \begin{bmatrix} -0.18 \\ 0.31 \\ 1.03 \\ 2.48 \\ 3.73 \end{bmatrix}$$

Thus,

$$\mathbf{v}^* = \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} \approx \begin{bmatrix} -0.40 \\ 0.35 \\ 16.1 \end{bmatrix},$$

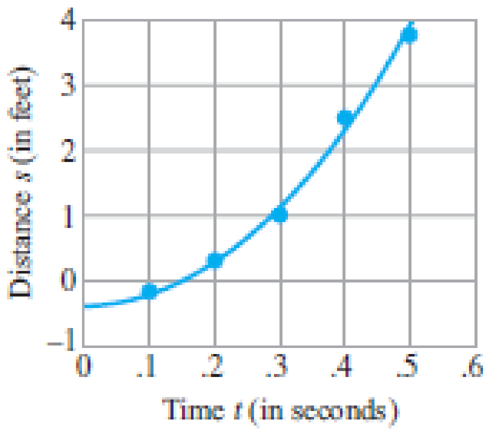
so the least squares quadratic fit is

$$s = -0.40 + 0.35t + 16.1t^2$$

From this equation we estimate that $\frac{1}{2}g = 16.1$ and hence that $g = 32.2$ ft/sec^2. Note that this equation also provides the following estimates of the initial displacement and velocity of the weight:

$$s_0 = a_0^* = -0.40 \text{ ft}$$
$$v_0 = a_1^* = \quad 0.35 \text{ ft/sec}$$

In Figure we have plotted the data points and the approximating polynomial.



3. Function Approximation; Fourier Series

In this subsection we will show how orthogonal projections can be used to approximate certain types of functions by simpler functions. The ideas explained here have important applications in engineering and science. Calculus is required.

All of the problems that we will study in this section will be special cases of the following general problem.

Approximation Problem: Given a function f that is continuous on an interval [*a*, *b*], find the "*best possible approximation*" to *f* using only functions from a specified subspace *W* of **C**[*a*, *b*].

Here are some examples of such problems:

(a) Find the best possible approximation to e^x over [0, 1] by a polynomial of the form $a_0 + a_1x + a_2x^2$.

(b) Find the best possible approximation to sin πx over [−1, 1] by a function of the form $a_0 + a_1e^x + a_2e^{2x} + a_3e^{3x}$.

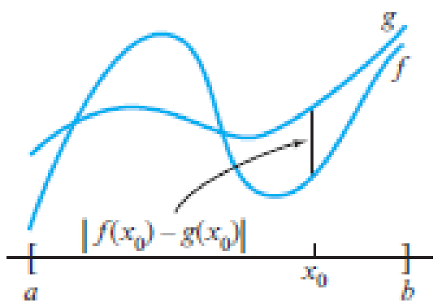
(c) Find the best possible approximation to x over [0, 2π] by a function of the form $a_0 + a_1 \sin x + a_2 \sin 2x + b_1 \cos x + b_2 \cos 2x$.

Notice that in the first example W is the subspace of **C**[0, 1] spanned by 1, x, and x^2; in the second example W is the subspace of **C**[−1, 1] spanned by 1, e^x, e^{2x}, and e^{3x}; and in the third example W is the subspace of **C**[0, 2π] spanned by 1, sin x, sin 2x, cos x, and cos 2x.

To solve *approximation problems* of the preceding types, we first need to make the phrase "*best approximation over* [a, b]" mathematically precise. To do this we will need some way of quantifying the error that results when one continuous function is approximated by another over an interval [a, b]. If we were to approximate $f(x)$ by $g(x)$, and if we were concerned only with the error in that approximation at a single point x_0, then it would be natural to define the error to be

$$\text{error} = |f(x_0) - g(x_0)|$$

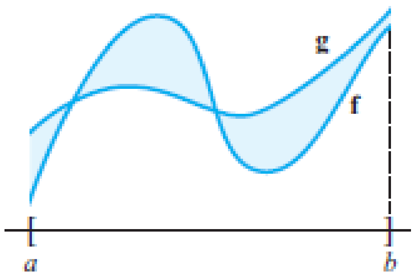
sometimes called the *deviation* between f and g at x_0 (Figure).



However, we are not concerned simply with measuring the error at a single point but rather with measuring it over the entire interval [a, b]. The difficulty is that an approximation may have small deviations in one part of the interval and large deviations in another. One possible way of accounting for this is to integrate the deviation $|f(x) - g(x)|$ over the interval [a, b] and define the error over the interval to be

$$\text{error} = \int_a^b |f(x) - g(x)|dx$$

Geometrically, this error is the area between the graphs of $f(x)$ and $g(x)$ over the interval [a, b] (Figure)—the greater the area, the greater the overall error.



Although the integral error is natural and appealing geometrically, most mathematicians and scientists generally favor the following alternative measure of error, called the mean square error:

$$\text{Mean square error} = \int_a^b [f(x) - g(x)]^2 dx$$

Mean square error emphasizes the effect of larger errors because of the squaring and has the added advantage that it allows us to bring to bear the theory of inner product spaces. To see how, suppose that f is a continuous function on [a, b] that we want to approximate by a function g from a subspace W of **C**[a, b], and suppose that **C**[a, b] is given the inner product:

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

It follows that

$$\|f-g\|^2 = \langle f-g, f-g\rangle = \int_a^b [f(x) - g(x)]^2 dx = \text{Mean square error}$$

so minimizing the mean square error is the same as minimizing $\|f-g\|^2$. Thus, the approximation problem posed informally at the beginning of this section can be restated more precisely as follows

Least Squares Approximation Problem

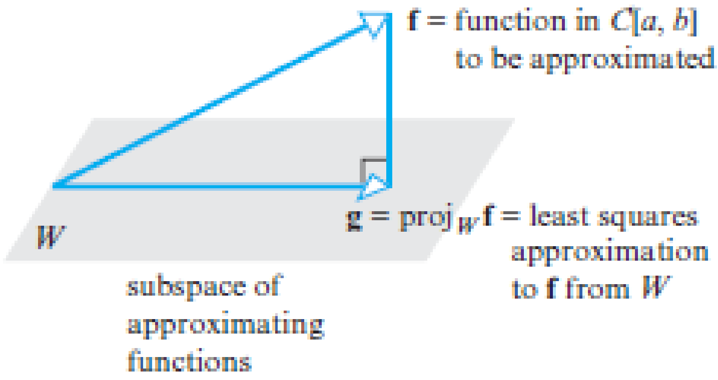
Let f be a function that is continuous on an interval [a, b], let **C**[a, b] have the inner product

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

and let W be a finite-dimensional subspace of **C**[a, b]. Find a function g in W that minimizes

$$\|f-g\|^2 = \int_a^b [f(x) - g(x)]^2 dx$$

Since $\|f - g\|^2$ and $\|f - g\|$ are minimized by the same function g, this problem is equivalent to looking for a function g in W that is closest to f. But we know from Theorem that g = **proj**(f, W) is such a function (Figure).



Thus, we have the following result.

Theorem: If f is a continuous function on [a, b], and W is a finite-dimensional subspace of **C**[a, b], then the function g in W that minimizes the mean square error

$$\int_a^b [f(x) - g(x)]^2 dx$$

is g = **proj**(f, W), where the orthogonal projection is relative to the inner product

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

The function g = **proj**(f, W) is called the *least squares approximation* to f from W.

Fourier Series

A function of the form

$$T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx \\ + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx$$

is called a *trigonometric polynomial*; if c_n and d_n are not both zero, then $T(x)$ is said to have order n. For example

$$T(x) = 2 + \cos x - 3\cos 2x + 7\sin 4x$$

is a trigonometric polynomial of order 4 with

$$c_0 = 2,\quad c_1 = 1,\quad c_2 = -3,\quad c_3 = 0,\quad c_4 = 0,\quad d_1 = 0,\quad d_2 = 0,\quad d_3 = 0,\quad d_4 = 7$$

It is evident from the general expression that the trigonometric polynomials of order n or less are the various possible linear combinations of

$$1,\quad \cos x,\quad \cos 2x, \ldots,\quad \cos nx,\qquad \sin x,\quad \sin 2x, \ldots,\quad \sin nx$$

It can be shown that these $2n + 1$ functions are linearly independent and thus form a basis for a $(2n + 1)$-dimensional subspace of **C**[a, b].

Let us now consider the problem of finding the *least squares approximation* of a continuous function $f(x)$ over the interval $[0, 2\pi]$ by a trigonometric polynomial of order n or less. As noted above, the least squares approximation to f from W is the *orthogonal projection* of f on W. To find this orthogonal projection, we must find an orthonormal basis $g_0, g_1, \ldots, g_{2n}$ for W, after which we can compute the orthogonal projection on W from the formula

$$\mathbf{proj}(f, W) = \langle f, g_0\rangle g_0 + \langle f, g_1\rangle g_1 + \cdots + \langle f, g_{2n}\rangle g_{2n}$$

Following the Theorem, an orthonormal basis for W can be obtained by applying the Gram–Schmidt process to the basis vectors in the system of trigonometric functions using the inner product

$$\langle f, g\rangle \int_0^{2\pi} f(x)g(x)dx$$

This yields the orthonormal basis

$$g_0 = \frac{1}{\sqrt{2\pi}}, \quad g_1 = \frac{1}{\sqrt{\pi}}\cos x, \ldots, \quad g_n = \frac{1}{\sqrt{\pi}}\cos nx,$$

$$g_{n+1} = \frac{1}{\sqrt{\pi}}\sin x, \ldots, \quad g_{2n} = \frac{1}{\sqrt{\pi}}\sin nx$$

If we introduce the notation:

$$a_0 = \frac{2}{\sqrt{2\pi}}\langle \mathbf{f}, g_0\rangle, \quad a_1 = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, g_1\rangle, \ldots, \quad a_n = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, g_n\rangle$$

$$b_1 = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, g_{n+1}\rangle, \ldots, \quad b_n = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, g_{2n}\rangle$$

Substituting these expressions into the orthogonal projection on W gives us

$$\mathbf{proj}(f, W) = \frac{a_0}{2} + \{a_1 \cos x + \cdots + a_n \cos nx\} + \{b_1 \sin x + \cdots + b_n \sin nx\}$$

where

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\, dx \text{ and } b_k = \frac{1}{\pi}\int_0^{2\pi} g(x)\sin kx\, dx$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the *Fourier coefficients* of $f(x)$.

Example (Least Squares Approximations) Find the least squares approximation of $f(x) = x$ on $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 2 or less;

(b) a trigonometric polynomial of order n or less.

(a)

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} x\,dx = 2\pi$$

For $k = 1, 2, \ldots$, integration by parts yields (verify)

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\cos kx\,dx = 0$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\sin kx\,dx = -\frac{2}{k}$$

Thus, the least squares approximation to x on [0, 2π] by a trigonometric polynomial of order 2 or less is

$$x \approx \frac{a_0}{2} + a_1\cos x + a_2\cos 2x + b_1\sin x + b_2\sin 2x$$

or

$$x \approx \pi - 2\sin x - \sin 2x$$

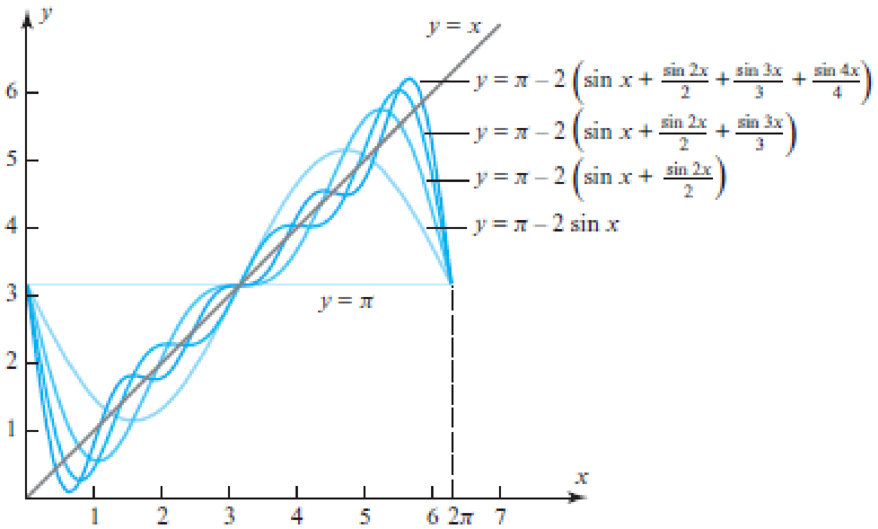
(b) The least squares approximation to x on [0, 2π] by a trigonometric polynomial of order n or less is

$$x \approx \frac{a_0}{2} + [a_1\cos x + \cdots + a_n\cos nx] + [b_1\sin x + \cdots + b_n\sin nx]$$

or, from (9a), (9b), and (9c),

$$x \approx \pi - 2\left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots + \frac{\sin nx}{n}\right)$$

The graphs of y = x and some of these approximations are shown in Figure.



The general form of the Fourier series for f over the interval [0, 2π] is

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx)$$