16

A Least Squares Model for Human Hearing: Fourier Series

The method of least squares approximation is motivated by energy considerations in a model for human hearing. Linear Algebra techniques are used to approximate a function over an interval by a linear combination of sine and cosine terms.

PREREQUISITES:	Inner product spaces Orthogonal projection $C[a, b]$ (the vector space of continuous functions on $[a, b]$)

INTRODUCTION

We begin with a brief discussion of the nature of sound and human hearing. Figure 16.1 is a schematic diagram of the ear showing its three main components: the outer ear, middle ear, and inner ear. Sound waves enter the outer ear where they are channeled to the eardrum, causing it to vibrate. Three tiny bones in the middle ear mechanically link the eardrum with the snail-shaped cochlea within the inner ear. These bones pass on the vibrations of the eardrum to a fluid within the cochlea. The cochlea contains thou-

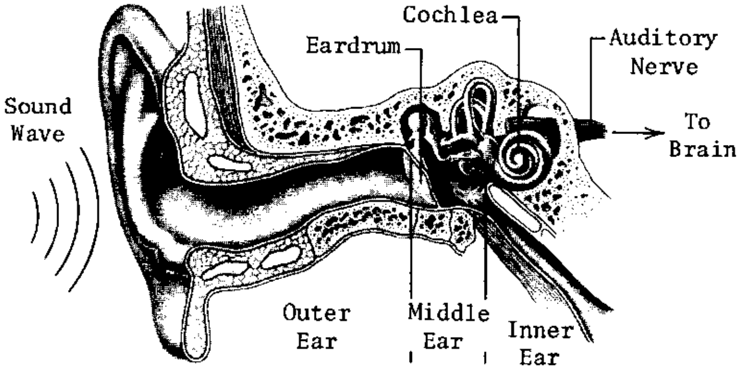

Figure 16.1

sands of minute hairs which oscillate with the fluid. Those near the entrance of the cochlea are stimulated by high frequencies and those near the tip are stimulated by low frequencies. The movements of these hairs activate nerve cells which send signals along various neural pathways to the brain, where the signals are interpreted as sound.

The sound waves themselves are variations in time of the air pressure. For the auditory system, the most elementary type of sound wave is a sinusoidal variation in the air pressure. This type of sound wave stimulates the hairs within the cochlea in such a way that a nerve impulse along a single neural pathway is produced (Fig. 16.2). A sinusoidal sound wave may be described by a function of time

$$q(t) = A_0 + A \sin \omega(t - \delta) \tag{16.1}$$

where $q(t)$ is the atmospheric pressure at the eardrum, A_0 is the normal atmospheric pressure, A is the maximum deviation of the pressure from the normal atmospheric pressure, $\omega/2\pi$ is the frequency of the wave in cycles per second, and δ is the phase angle of the wave. To be perceived as sound, such sinusoidal waves must have frequencies within a certain range. For humans this range is roughly 20 cps to 20,000 cps. Frequencies outside of this range will not stimulate the hairs within the cochlea enough to produce nerve signals.

To a reasonable degree of accuracy, the ear is a linear system. This means that if a complex sound wave is a finite sum of sinusoidal components of different amplitudes, frequencies, and phase angles, say

$$q(t) = A_0 + A_1 \sin \omega_1(t - \delta_1)$$
$$+ A_2 \sin \omega_2(t - \delta_2) + \cdots + A_n \sin \omega_n(t - \delta_n), \tag{16.2}$$

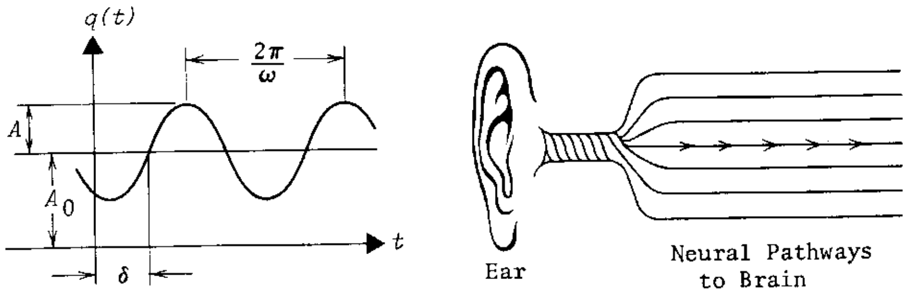

Figure 16.2

then the response of the ear consists of nerve impulses along the same neural pathways that would be stimulated by the individual components (Fig. 16.3).

Let us now consider some periodic sound wave $p(t)$ with period T (i.e., $p(t) \equiv p(t+T)$) which is *not* a finite sum of sinusoidal waves. If we examine the response of the ear to such a periodic wave, we find that it is the same as the response to some wave which is the sum of sinusoidal waves. That is, there is some sound wave $q(t)$ as given by Eq. (16.2) which produces the same response as $p(t)$, even though $p(t)$ and $q(t)$ are different functions of time.

We now want to determine the frequencies, amplitudes, and phase angles of the sinusoidal components of $q(t)$. Since $q(t)$ produces the same response as the periodic wave $p(t)$, it is reasonable to expect that $q(t)$ has the same period T as $p(t)$. This requires that each sinusoidal term in $q(t)$ have period T. Consequently, the frequencies of the sinusoidal components must be integer multiples of

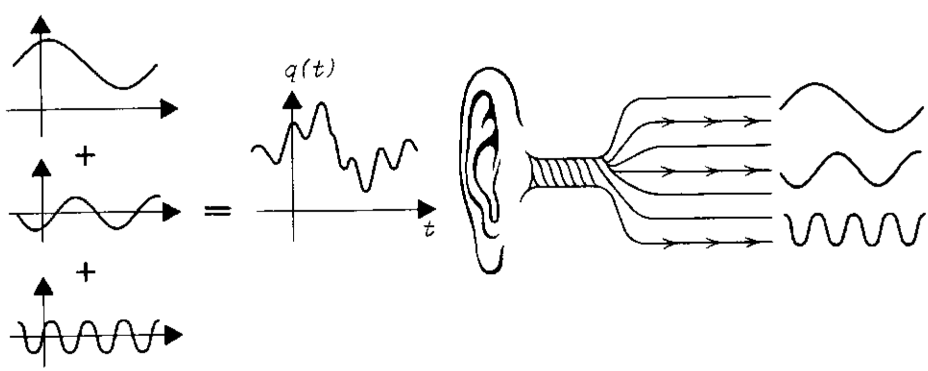

Figure 16.3

the basic frequency $1/T$ of the $p(t)$. That is, the ω_k in Eq. (16.2) must be of the form

$$\omega_k = 2k\pi/T, \qquad k = 1, 2, \ldots$$

But since the ear cannot perceive sinusoidal waves with frequencies greater than 20,000 cps, we may omit those values of k for which $\omega_k/2\pi = k/T$ is greater than 20,000. Thus, $q(t)$ is of the form

$$q(t) = A_0 + A_1 \sin\frac{2\pi}{T}(t - \delta_1) + \cdots + A_n \sin\frac{2n\pi}{T}(t - \delta_n) \qquad (16.3)$$

where n is the largest integer such that n/T is not greater than 20,000.

We now turn our attention to the values of the amplitudes A_0, $A_1, \ldots, A_n$ and the phase angles $\delta_1, \delta_2, \ldots, \delta_n$ which appear in Eq. (12.3). There is some criterion by which the auditory system "picks" these values so that $q(t)$ produces the same response as $p(t)$. To examine this criterion, let us set

$$e(t) = p(t) - q(t).$$

If we consider $q(t)$ as an approximation to $p(t)$, then $e(t)$ is the error in this approximation; an error which the ear cannot perceive. In terms of $e(t)$, the criterion for the determination of the amplitudes and the phase angles is that the quantity

$$\int_0^T [e(t)]^2 dt$$

be as small as possible. We cannot go into the physiological reasons for this, but we can remark that this expression is proportional to the *acoustic energy* of the error wave $e(t)$ over one period. In other words, it is the energy of the difference between the two sound waves $p(t)$ and $q(t)$ which determine whether the ear perceives any difference between them. If this energy is as small as possible, then the two waves produce the same sensation of sound.

In general, suppose we want to approximate a function $p(t)$ by another function $q(t)$ from a certain class. If the criterion of approximation is the smallness of the integral

$$\int_0^T [p(t) - q(t)]^2 dt, \qquad (16.4)$$

then we call $q(t)$ the *least squares approximation* to $p(t)$ over the interval $[0, T]$. The integral in (16.4) is called the *least squares error*, or the *mean squares error*, of the approximation. This criterion arises quite naturally in a wide variety of engineering and scientific approximation problems. Besides the acoustics problem just discussed, some other examples are

1. Let $T(x)$ be the temperature distribution in a uniform rod lying along the x-axis from $x = 0$ to $x = \ell$ (Fig. 16.4). The thermal energy in the rod is proportional to the integral

$$\int_0^{\ell} [T(x)]^2 dx.$$

The closeness of an approximation $q(x)$ to $T(x)$ can be judged according to the thermal energy of the difference of the two temperature distributions. That energy is proportional to

$$\int_0^{\ell} [T(x) - q(x)]^2 dx,$$

which is a least squares criterion.

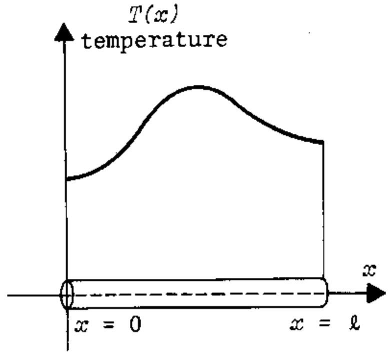

Figure 16.4

2. Let $E(t)$ be a periodic voltage across a resistor in an electrical circuit. The electrical energy transferred to the resistor during one period T is proportional to

$$\int_0^{T} [E(t)]^2 dt.$$

If $q(t)$ has the same period as $E(t)$ and is to be an approximation to $E(t)$, then the criterion

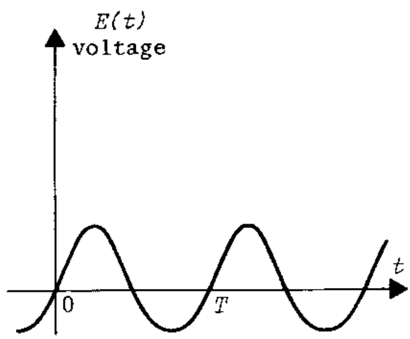

Figure 16.5

of closeness might be taken as the energy of the difference voltage. This is proportional to

$$\int_0^T [E(t) - q(t)]^2 dt,$$

which is again a least squares criterion.

3. Let $y(x)$ be the vertical displacement of a uniform flexible string whose equilibrium position is along the x-axis from $x = 0$ to $x = \ell$. The elastic potential energy of the string is proportional to

$$\int_0^\ell [y(x)]^2 dx.$$

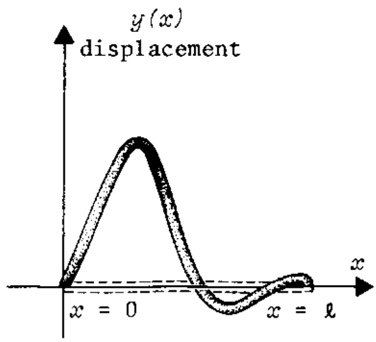

Figure 16.6

If $q(x)$ is to be an approximation to the displacement, then as before, the energy integral

$$\int_0^\ell [y(x) - q(x)]^2 dx$$

determines a least squares criterion for the closeness of the approximation.

Least squares approximation is also used in situations where there is no a priori justification for its use, such as for approximating business cycles, population growth curves, sales curves, and so forth. It is used in these cases because of its mathematical simplicity. In general, if no other error criterion is immediately apparent for an approximation problem, the least squares criterion is the one most often chosen.

In the next section we develop the mathematical theory of the least squares approximation of a function by a linear combination of sinusoidal functions.

General Theory

Let $f(t)$ be a given continuous function defined over an interval of the t-axis. We first consider the case when the interval is

$[0, 2\pi]$ and later consider the general case when the interval is $[0, T]$ for arbitrary T. Analogous to Eq. (16.3) with $T = 2\pi$, we desire to approximate $f(t)$ by a function of the form

$$g(t) = A_0 + A_1 \sin(t - \delta_1) + A_2 \sin 2(t - \delta_2) + \cdots + A_n \sin n(t - \delta_n)$$

for some fixed integer n. Since $\sin k(t - \delta_k)$ is expressible as a linear combination of $\sin kt$ and $\cos kt$, we can write $g(t)$ in the alternate form

$$\begin{aligned} g(t) = c_0 + c_1 \cos t + c_2 \cos 2t + \cdots + c_n \cos nt \\ + d_1 \sin t + d_2 \sin 2t + \cdots + d_n \sin nt\,. \end{aligned} \tag{16.5}$$

Such a function is called a *trigonometric polynomial of order n*. Our problem is to find values of $c_0, c_1, \ldots, c_n, d_1, \ldots, d_n$ such that $g(t)$ is the least squares approximation to $f(t)$ over the interval $[0, 2\pi]$. That is, the coefficients are to be chosen so that the least squares error

$$\int_0^{2\pi} [f(t) - g(t)]^2 dt \tag{16.6}$$

is as small as possible.

Since the integral in Eq. (16.6) is a function of the $2n + 1$ coefficients $c_0, c_1, \ldots, c_n, d_1, \ldots, d_n$, it is possible to use calculus to find the minimum value of the least squares error and the corresponding values of these $2n + 1$ coefficients. However, an approach using Linear Algebra will give us greater insight into the nature of the approximation process. Moreover, the method we discuss can be applied to many least square problems besides those in this text. We will need the following three facts:

1. The function $f(t)$ we are attempting to approximate may be viewed as a vector in the vector space $C[0, 2\pi]$ — the space of all continuous functions on $[0, 2\pi]$.

2. Since the approximating function $g(t)$ is a linear combination of $1, \cos t, \ldots, \cos nt, \sin t, \ldots, \sin nt$, we may view $g(t)$ as a vector in the subspace W of $C[0, 2\pi]$ spanned by these $2n + 1$ vectors.

3. Since

$$\|f - g\| = \sqrt{\int_0^{2\pi} [f(t) - g(t)]^2 dt} \tag{16.7}$$

is the distance between $f(t)$ and $g(t)$ in the norm generated by the inner product

$$<u, v> = \int_0^{2\pi} u(t)v(t)\, dt, \tag{16.8}$$

the least squares error

$$\int_0^{2\pi} [f(t) - g(t)]^2 dt \tag{16.9}$$

represents the square of the distance $\|f - g\|$.

In light of these remarks, the problem of finding a trigonometric polynomial $g(t)$ which minimizes the least squares error given by (16.9) is equivalent to the problem of finding a vector g in the subspace W which minimizes the distance $\|f - g\|$. The latter problem can be solved by use of the following theorem from the theory of inner product spaces (Fig. 16.7).

THEOREM 16.1 *Let f be a vector in an inner product space and let W be a finite-dimensional subspace. Then the vector g in W which minimizes the distance $\|f - g\|$ is* $\text{proj}_W f$, *the orthogonal projection of f onto W. If the vectors*

$$g_0, g_1, \ldots, g_m$$

form an orthonormal basis for W, then $\text{proj}_W f$ *is gvven by*

$$\text{proj}_W f = <f, g_0>g_0 + <f, g_1>g_1 + \cdots + <f, g_m>g_m \tag{16.10}$$

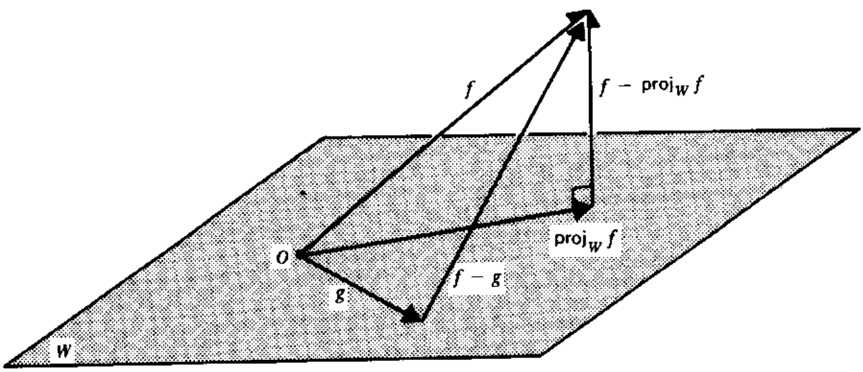

Figure 16.7

To apply this theorem, we must first find an orthonormal basis for the subspace W spanned by the $2n+1$ vectors 1, $\cos t$, $\cos 2t, \ldots,$ $\cos nt$, $\sin t$, $\sin 2t, \ldots, \sin nt$. A direct calculation (see Exercise 16.6) verifies that these $2n+1$ vectors are orthogonal relative to the inner product (16.8). Consequently, we need only divide each one of these vectors by its length to generate an orthonormal basis for W. The result is (see Exercise 16.7):

$$g_0 = \frac{1}{\sqrt{2\pi}},\ g_1 = \frac{1}{\sqrt{\pi}} \cos t,\ \ldots,\ g_n = \frac{1}{\sqrt{\pi}} \cos nt,$$

$$g_{n+1} = \frac{1}{\sqrt{\pi}} \sin t,\ \ldots,\ g_{2n} = \frac{1}{\sqrt{\pi}} \sin nt.$$

The orthogonal projection of f onto W is then given by (16.10):

$$\text{proj}_W f = \langle f, g_0\rangle \frac{1}{\sqrt{2\pi}} + \langle f, g_1\rangle \frac{1}{\sqrt{\pi}} \cos t + \cdots + \langle f, g_n\rangle \frac{1}{\sqrt{\pi}} \cos nt$$

$$+ \langle f, g_{n+1}\rangle \frac{1}{\sqrt{\pi}} \sin t + \cdots + \langle f, g_{2n}\rangle \frac{1}{\sqrt{\pi}} \sin nt. \quad (16.11)$$

To simplify our notation, let us define

$$a_0 = 2\langle f, g_0\rangle \frac{1}{\sqrt{2\pi}} = \frac{1}{\pi}\int_0^{2\pi} f(t)\,dt\,,$$

$$a_k = \langle f, g_k\rangle \frac{1}{\sqrt{\pi}} = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\,dt, \qquad k = 1, 2, \ldots, n$$

$$b_k = \langle f, g_{n+k}\rangle \frac{1}{\sqrt{\pi}} = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\,dt, \qquad k = 1, 2, \ldots, n.$$

Equation (16.11) can then be written as

$$\text{proj}_W f = \tfrac{1}{2}a_0 + a_1\cos t + \cdots + a_n\cos nt + b_1\sin t + \cdots + b_n\sin nt.$$

In summary, we have the following result:

THEOREM 16.2 *If $f(t)$ is continuous on $[0, 2\pi]$, the trigonometric function $g(t)$ of the form*

$$g(t) = \tfrac{1}{2}a_0 + a_1\cos t + \cdots + a_n\cos nt + b_1\sin t + \cdots + b_n\sin nt$$

which minimizes the least squares error

$$\int_0^{2\pi} [f(t) - g(t)]^2\,dt$$

has coefficients

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\,dt, \qquad k = 0, 1, 2, \ldots, n$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\,dt, \qquad k = 1, 2, \ldots, n.$$

If the original function $f(t)$ is defined over the interval $[0, T]$ instead of $[0, 2\pi]$, a change of scale will yield the following result (see Exercise 16.8):

THEOREM 16.3 *If $f(t)$ is continuous on $[0, T]$, the trigonometric function $g(t)$ of the form*

$$g(t) = \tfrac{1}{2}a_0 + a_1\cos\frac{2\pi}{T}t + \cdots + a_n\cos\frac{2n\pi}{T}t$$

$$+ b_1\sin\frac{2\pi}{T}t + \cdots + b_n\sin\frac{2n\pi}{T}t$$

which minimizes the least squares error

$$\int_0^T [f(t) - g(t)]^2 dt$$

has coefficients

$$a_k = \frac{2}{T}\int_0^T f(t)\cos\frac{2k\pi t}{T}\,dt, \qquad k = 0, 1, 2, \ldots, n$$

$$b_k = \frac{2}{T}\int_0^T f(t)\sin\frac{2k\pi t}{T}\,dt, \qquad k = 1, 2, \ldots, n.$$

EXAMPLE 16.1 Let a sound wave $p(t)$ have a saw-tooth pattern with a basic frequency of 5000 cps (Fig. 16.8). Units have been chosen so that the normal atmospheric pressure is at the zero level, and the maximum amplitude of the wave is A. The basic period of the wave is $T = 1/5000 = .0002$ seconds. From $t = 0$ to $t = T$, $p(t)$ has the equation

$$p(t) = \frac{2A}{T}\left(\frac{T}{2} - t\right).$$

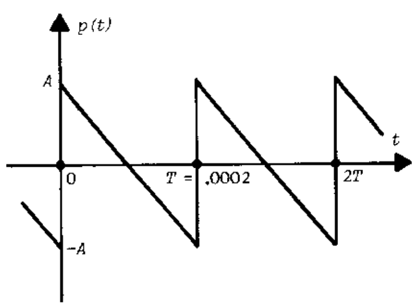

Figure 16.8

Theorem 16.3 then yields the following (verify):

$$a_0 = \frac{2}{T}\int_0^T p(t)\,dt = \frac{2}{T}\int_0^T \frac{2A}{T}\left(\frac{T}{2} - t\right)dt = 0,$$

$$a_k = \frac{2}{T}\int_0^T p(t)\cos\frac{2k\pi t}{T}\,dt = \frac{2}{T}\int_0^T \frac{2A}{T}\left(\frac{T}{2} - t\right)\cos\frac{2k\pi t}{T}\,dt = 0, \quad k = 1, 2, \ldots$$

$$b_k = \frac{2}{T}\int_0^T p(t)\sin\frac{2k\pi t}{T}\,dt = \frac{2}{T}\int_0^T \frac{2A}{T}\left(\frac{T}{2} - t\right)\sin\frac{2k\pi t}{T}\,dt = \frac{2A}{k\pi}, \quad k = 1, 2, \ldots$$

Let us investigate how the sound wave $p(t)$ is perceived by the human ear. We notice that $4/T = 20{,}000$ cps, so that we need only go up to $k = 4$ in the above formulas. The least squares approximation to $p(t)$ is then

$$q(t) = \frac{2A}{\pi}\left[\sin\frac{2\pi}{T}t + \frac{1}{2}\sin\frac{4\pi}{T}t + \frac{1}{3}\sin\frac{6\pi}{T}t + \frac{1}{4}\sin\frac{8\pi}{T}t\right].$$

The four sinusoidal terms have frequencies of 5,000, 10,000, 15,000, and 20,000 cps, respectively. In Fig. 16.9 we have plotted $p(t)$ and $q(t)$ over one period. Although $q(t)$ is not a very good point-by-point approximation to $p(t)$, to the ear, both $p(t)$ and $q(t)$ produce the same sensation of sound.

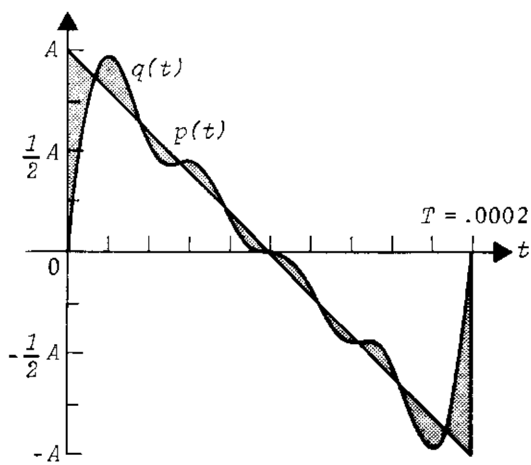

Figure 16.9

As might be expected, the least squares approximation becomes better as the number of terms in the approximating trigonometric polynomial becomes larger. In more advanced courses, it is shown that the least squares error tends to zero as n approaches infinity. For a function $f(t)$ defined over the interval $[0, 2\pi]$, this limiting approximation is denoted by

$$f(t) = \tfrac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \sin kt) \qquad (16.12)$$

and is called the *Fourier series* of $f(t)$ on the interval $[0, 2\pi]$. The equality in this equation denotes an equality between the two sides of the equation considered as vectors in $C[0, 2\pi]$. To be precise, Eq. (16.12) denotes the fact that the quantity

$$\int_0^{2\pi}\left[f(t) - \tfrac{1}{2}a_0 - \sum_{k=1}^{n}(a_k \cos kt + b_k \sin kt)\right]^2 dt$$

tends to zero as n approaches infinity. Whether the Fourier series of $f(t)$ converges to $f(t)$ for each t is another question, and a more difficult one. For most continuous functions encountered in applications, the Fourier series does indeed converge to its corresponding function for each value of t.

EXERCISES

16.1 Find the trigonometric polynomial of order three which is the least squares approximation to the function $f(t) = (t - \pi)^2$ over the interval $[0, 2\pi]$.

16.2 Find the trigonometric polynomial of order four which is the least squares approximation to the function $f(t) = t^2$ over the interval $[0, T]$.

16.3 Find the trigonometric polynomial of order four which is the least squares approximation to the function $f(t)$ over the interval $[0, 2\pi]$ where

$$f(t) = \begin{cases} \sin t & 0 \le t \le \pi \\ 0 & \pi < t \le 2\pi \end{cases}.$$

16.4 Find the trigonometric polynomial of arbitrary order n which is the least squares approximation to the function $f(t) = \sin \frac{1}{2} t$ over the interval $[0, 2\pi]$.

16.5 Find the trigonometric polynomial of arbitrary order n which is the least squares approximation to the function $f(t)$ over the interval $[0, T]$ where

$$f(t) = \begin{cases} t & 0 \leq t \leq \frac{1}{2}T \\ T - t & \frac{1}{2}T < t \leq T \end{cases}.$$

16.6 Show that the $2n + 1$ functions

$$1, \cos t, \cos 2t, \ldots, \cos nt, \sin t, \sin 2t, \ldots, \sin nt$$

are orthogonal over the interval $[0, 2\pi]$ relative to the inner product $\langle u, v\rangle$ defined by Eq. (16.8).

16.7 For the distance formula defined in (16.7), show that

(a) $\|1\| = \sqrt{2\pi}$

(b) $\|\cos kt\| = \sqrt{\pi}$ for $k = 1, 2, \ldots$

(c) $\|\sin kt\| = \sqrt{\pi}$ for $k = 1, 2, \ldots$

16.8 If $f(t)$ is defined and continuous on the interval $[0, T]$, show that $f(T\tau/2\pi)$ is defined and continuous for τ in the interval $[0, 2\pi]$. Use this fact to show how Theorem 16.3 follows from Theorem 16.2.