A Historical Perspective of Spectrum Estimation

ENDERS A. ROBINSON

Invited Paper

Abstract—The prehistory of spectral estimation has its roots in ancient times with the development of the calendar and the clock. The work of Pythagoras in 600 B.C. on the laws of musical harmony found mathematical expression in the eighteenth century in terms of the wave equation. The struggle to understand the solution of the wave equation was finally resolved by Jean Baptiste Joseph de Fourier in 1807 with his introduction of the Fourier series. The Fourier theory was extended to the case of arbitrary orthogonal functions by Sturm and Liouville in 1836. The Sturm–Liouville theory led to the greatest empirical success of spectral analysis yet obtained, namely the formulation of quantum mechanics as given by Heisenberg and Schrödinger in 1925 and 1926. In 1929 John von Neumann put the spectral theory of the atom on a firm mathematical foundation in his spectral representation theorm in Hilbert space. Meanwhile, Wiener developed the mathematical theory of Brownian movement in 1923, and in 1930 he introduced generalized harmonic analysis, that is, the spectral representation of a stationary random process. The common ground of the spectral representations of von Neumann and Wiener is the Hilbert space; the von Neumann result is for a Hermitian operator, whereas the Wiener result is for a unitary operator. Thus these two spectral representations are related by the Cayley–Möbius transformation. In 1942 Wiener applied his methods to problems of prediction and filtering, and his work was interpreted and extended by Norman Levinson. Wiener in his empirical work put more emphasis on the autocorrelation function than on the power spectrum.

The modern history of spectral estimation begins with the breakthrough of J. W. Tukey in 1949, which is the statistical counterpart of the breakthrough of Fourier 142 years earlier. This result made possible an active development of empirical spectral analysis by research workers in all scientific disciplines. However, spectral analysis was computationally expensive. A major computational breakthrough occurred with the publication in 1965 of the fast Fourier transform algorithm by J. S. Cooley and J. W. Tukey. The Cooley–Tukey method made it practical to do signal processing on waveforms in either the time or the frequency domain, something never practical with continuous systems. The Fourier transform became not just a theoretical description, but a tool. With the development of the fast Fourier transform the field of empirical spectral analysis grew from obscurity to importance, and is now a major discipline. Further important contributions were the introduction of maximum entropy spectral analysis by John Burg in 1967, the development of spectral windows by Emmanuel Parzen and others starting in the 1950's, the statistical work of Maurice Priestley and his school, hypothesis testing in time series analysis by Peter Whittle starting in 1951, the Box–Jenkins approach by George Box and G. M. Jenkins in 1970, and autoregressive spectral estimation and order-determining criteria by E. Parzen and H. Akaike starting in the 1960's. To these statistical contributions must be added the equally important engineering contributions to empirical spectrum analysis, which are not treated at all in this paper, but form the subject matter of the other papers in this special issue.

I. Introduction

SPECTRAL estimation has its roots in ancient times, with the determination of the length of the day, the phases of the moon, and the length of the year. The calendar and the clock resulted from empirical spectral analysis. In modern times, credit for the empirical discovery of spectra goes to the diversified genius of Sir Isaac Newton [1]. But the great interest in spectral analysis made its appearance only a little more than a century ago. The prominent German chemist Robert Wilhelm Bunsen (1811–1899) repeated Newton's experiment of the glass prism. Only Bunsen did not use the sun's rays as Newton did. Newton had found that a ray of sunlight is expanded into a band of many colors, the spectrum of the rainbow. In Bunsen's experiment, the role of pure sunlight was replaced by the burning of an old rag that had been soaked in a salt solution (sodium chloride). The beautiful rainbow of Newton did not appear. The spectrum, which Bunsen saw, only exhibited a few narrow lines, nothing more. One of the lines was a bright yellow.

Bunsen conveyed this result to Gustav Robert Kirchhoff (1824–1887), another well-known German scientist. They knew that the role of the glass prism consisted only in sorting the incident rays of light into their respective wavelengths (the process known as dispersion). The Newton rainbow was the extended continuous band of the solar spectrum; it indicates that all wavelengths of visible light are present in pure sunlight. The yellow line, which appeared when the light source was a burning rag, indicated that the spectrum of table salt contained a single specific wavelength. Further experiments showed that this yellow line belonged to the element sodium. No matter what the substance in which sodium appeared, that element made its whereabouts known by its bright yellow spectral line. As time went on, it was found that every chemical element has its own characteristic spectrum, and that the spectrum of a given element is always the same, no matter in what compound or substance the element is found. Thus the spectrum identifies the element, and in this way we can tell what elements are in substances from the distant stars to microscopic objects.

The successes of spectral analysis were colossal. However, the spectral theory of the elements could not be explained by classical physics. As we know, quantum physics was born and spectral theory was explained in 1925 and 1926 by the work of Werner Heisenberg (1901–1976) and Erwin Schrödinger (1887–1961). In this paper, we will show how spectral theory developed in the path to this great achievement.

Although most of the glamour of spectral theory has been associated with quantum physics, we will not neglect the parallel path taken in classical physics. Although the two paths began diverging with the work of Charles Sturm (1803–1855) and Joseph Liouville (1809–1882) on the spectral theory of differential equations, we will see that the final results, namely, the spectral representation of John von Neumann (1903–1957) for quantum physics, and that of Norbert Wiener (1894–1964) for classical physics, are intimately related.

Because light has high frequencies, our instruments cannot respond fast enough to directly measure the waveforms. Instead, the instruments measure the amount of energy in the

Manuscript received January 22, 1982; revised April 28, 1982. This paper is dedicated to Norman Levinson, August 11, 1912–October 10, 1975, my teacher in differential equations.

The author is with the Department of Theoretical and Applied Mechanics and the Department of Geological Sciences, Cornell University, Ithaca, NY 14853.

0018-9219/82/0900-0885$00.75

frequency bands of interest. The measurement and analysis of the spectra of other types of signals, however, take different forms. With lower frequency signals, such as mechanical vibrations, speech, sonar signals, seismic traces, cardiograms, stock market data, and so on, we can measure the signals as functions of time (that is, as time series) and then find the spectra by computation. With the advent of the digital computer, numerical spectrum estimation has become an important field of study.

Let us say a few words about the terms "spectrum" and "spectral." Sir Isaac Newton introduced the scientific term "spectrum" using the Latin word for an image. Today, in English, we have the word *specter* meaning ghost or apparition, and the corresponding adjective *spectral*. We also have the scientific word *spectrum* and the dictionary lists the word *spectral* as the corresponding adjective. Thus "spectral" has two meanings. Many feel that we should be careful to use "spectrum" in place of "spectral" a) whenever the reference is to data or physical phenomena, and b) whenever the word modifies "estimation." They feel that the word "spectral," with its unnecessary ghostly interpretations, should be confined to those usages in a mathematical discipline where the term is deeply embedded.

The material in this paper through Section XII surveying the period from antiquity through Levinson and Wiener can be described as "The Prehistory of Spectrum Estimation" to emphasize that spectrum estimation is interpreted as estimation from data. The remaining sections may be described as "Some Pioneering Contributions to the Development of Methods of Spectrum Estimation."

Modern spectrum estimation began with the breakthrough for the analysis of short time series made by J. W. Tukey in 1949. This work led to a great blossoming forth of spectrum analysis techniques. Despite the advances in digital computing machinery, such computations were still expensive. The next great breakthrough occurred with the discovery of the fast Fourier transform in 1965 independently by J. W. Cooley and J. W. Tukey and by Gordon Sande. This development, in conjunction with silicon chip technology, has brought spectrum analysis to bear on a wide range of problems. Another breakthrough occurred with the introduction of maximum-entropy methods into spectrum analysis by John Burg in 1967.

II. Taylor Series

At the time when calculus was introduced in the seventeenth century by Newton and Leibnitz, the concept of a mathematical "function" entailed restricted properties, which in the course of time were gradually made less severe. In those days, the observations of natural events seemed to indicate that continuous relations always existed between physical variables. This view was reinforced by the formulation of the laws of nature on the basis of differential equations, as exemplified by Newton's laws. Thus it became commonplace to assume that any function describing physical phenomena would be differentiable. The idea of a function that changes in some capricious or random way, and thus does not allow any analytic formula for its representation, did not enter into the thinking of the mathematicians of that time. It was thus very natural for Brook Taylor (1685–1731) [2], a contemporary of Newton, to introduce the concept of "analytic function." The Taylor series expands an analytic function as an infinite summation of component functions. More precisely, the Taylor series expands a function $f(x)$, which is analytic in the neighborhood of a certain point $x = a$, into an infinite series whose coefficients are the successive derivatives of the function at the given point

$$f(a+h) = f(a) + \frac{1}{1!} f'(a)\, h + \frac{1}{2!} f''(a)\, h^2 + \cdots.$$

Thus analytic functions are functions which can be differentiated to any degree. We know that the definition of the derivative of any order at the point $x = a$ does not require more than knowledge of the function in an arbitrarily small neighborhood of the point $x = a$. The astonishing property of the Taylor series is that the shape of the function at a finite distance h from the point $x = a$ is uniquely determined by the behavior of the function in the infinitesimal vicinity of the point $x = a$. Thus the Taylor series implies that an analytic function has a very strong interconnected structure; by studying the function in a small vicinity of the point $x = a$, we can precisely predict what happens at the point $x = a + h$, which is at a finite distance from the point of study. This property, however, is restricted to the class of analytic functions. The best known analytic functions are, of course, the sine and cosine functions, the polynomials, and the rational functions (away from their poles).

III. The Daniel Bernoulli Solution of the Wave Equation

The great Greek mathematician Pythagoras (ca 600 B.C.) was the first to consider a purely physical problem in which spectrum analysis made its appearance. Pythagoras studied the laws of musical harmony by generating pure sine vibrations on a vibrating string, fixed at its two endpoints. This problem excited scientists since ancient days, but the mathematical turning point came in the eighteenth century when it was recognized that the vertical displacement $u(x, t)$ of the vibrating string satisfies the wave equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0.$$

Here x is the horizontal coordinate and t is the time. The constant c is a physical quantity characteristic of the material of the string, and represents the velocity of the traveling waves on the string. Because the endpoints $x = 0$ and $x = \pi$ are fixed, we have the boundary conditions

$$u(0, t) = u(\pi, t) = 0.$$

(Note that for simplicity, we have taken the string to be of length π.) The problem of constructing the solution of the wave equation was then attacked by some of the greatest mathematicians of all time, and in so doing, they paved the way for the theory of spectrum analysis.

One of the finest results was that of Daniel Bernoulli (1700–1782) [3] in 1738. He introduced the method of separation of variables in which a trial solution is constructed as the product of a function of x alone, and a function of t alone. Thus, he wrote

$$u(x, t) = X(x)\, T(t).$$

Putting this trial solution in the differential equation and solving, he found the solutions

$$\cos kx \cos kct, \qquad \cos kx \sin kct$$

$$\sin kx \cos kct, \qquad \sin kx \sin kct.$$

However, the boundary condition at $x = 0$ excludes the solutions involving $\cos kx$, and so the possible solutions are reduced to the two choices

$$\sin kx \cos kct, \qquad \sin kx \sin kct.$$

The boundary condition at $x = \pi$ requires that the value of k to be an integer. In view of the linearity of the wave equation, any superposition of solutions gives a solution. Bernoulli thus gave the following solution:

$$u(x, t) = \sum_{k=1}^{\infty} \sin kx (A_k \cos kct + B_k \sin kct)$$

where the A_k and B_k are arbitrary constants. Bernoulli made the claim that this infinite sum is the general solution of the equation for the vibrating string. The implications of Bernoulli's claim were startling. From the principles of mechanics it was known that the initial displacement and initial velocity of the string could be prescribed in an arbitrary way. That is, it was known that at the initial time $t = 0$, both $u(x, 0)$ and $\dot{u}(x, 0)$ could have any functional form. (Note that the dot over a function indicates differentiation with respect to time, so $\dot{u}$ represents the velocity of the string in the vertical direction.) However, Bernoulli's solution gives explicit expressions for initial displacement and initial velocity, namely,

$$u(x, 0) = \sum_{k=1}^{\infty} A_k \sin kx$$

$$\dot{u}(x, 0) = c \sum_{k=1}^{\infty} kB_k \sin kx.$$

Thus Bernoulli's solution implied that each of two arbitrary functions $u(x, 0)$ and $\dot{u}(x, 0)$ could be expanded in the interval $0 \leqslant x \leqslant \pi$ in the form of an infinite series of sine functions. However, this result could not be explicitly demonstrated in Bernoulli's time.

Bernoulli's result can be expressed in the following way. Let the initial displacement $u(x, 0)$ be an arbitrary nonanalytic function $f(x)$. Then we have the expansion

$$f(x) = \sum_{k=0}^{\infty} A_k \sin kx$$

which says that a nonanalytic function $f(x)$ can be expressed as an infinite summation of analytic functions $\sin kx$ with weighting coefficients A_k. This result was a paradox at the time, and it led to a historical controversy of whether the function $f(x)$ could be freely chosen or must be restricted to the class of analytic functions. From the physical point of view, $f(x)$, which is the initial displacement of the string, could be freely chosen. From the then contemporary mathematical point of view, $f(x)$, which is an infinite summation of analytic functions, must be analytic. This view was believed by all the eminent mathematicians of the day.

Two of the greatest mathematicians who ever lived then set out to find the coefficients A_k of this expansion. Multiply each side by $\sin nx$ and integrate between 0 and π. Because

$$\int_0^{\pi} \sin kx \sin nx \, dx = \begin{cases} \pi/2, & \text{when } k = n \\ 0, & \text{when } k \neq n \end{cases}$$

the result found by L. Euler (1707–1783) [4] and J. L. Lagrange (1736–1813) [5] is

$$A_n = \frac{2}{\pi} \int_0^{\pi} f(x) \sin nx \, dx.$$

This is the point at which the question stood at the start of the nineteenth century.

IV. Jean Baptiste Joseph de Fourier and the Sinusoidal Spectral Theory

On December 21, 1807 the engineer Jean Baptiste Joseph de Fourier (1768–1830) [6] addressed the French Academy and made a claim that appeared incredible to the eminent mathematicians who were members of the Academy. As it turned out, one of the greatest advances in the history of mathematics, an innovation which was to occupy much of the attention of the mathematical community for over a century, was made by an engineer. Fourier said at that historic meeting that an arbitrary function, defined over a finite interval by any rough and even discontinuous graph, could be represented as an infinite summation of cosine and sine functions. The distinguished and brilliant academicians questioned the validity of Fourier's theorem, for they believed that any superposition of cosine and sine functions could only give an analytic function, that is, an infinitely differentiable function. An analytic function, of course, could never be discontinuous, and thus was very far removed from some arbitrarily drawn graph. In fact, Taylor's theorem stated that an analytic function had the property that, given its shape in an infinitesimal interval, the continuation of its course to the right and left by finite amounts was uniquely determined (the so-called process of analytic continuation). The academicians and the other great mathematicians of the time could not reconcile the property of analytic continuation with Fourier's theorem. How could the physical reasoning of an engineer stand up against the weight of the analytic reasoning of some of the most eminent mathematicians of all time? These were the days when many great men were at the peak of their powers. Yet Fourier stood alone in defending his theorem.

As we have seen, the concept of analytic function requires a strong interconnection of the values of a function, where knowledge at one point allows us to predict the value at a point at a finite distance h. This prediction mechanism is embodied in the Taylor series expansion. However, a nonanalytic function, such as a rough and discontinuous function, does not demand any such prediction mechanism between the immediate vicinity of a point and its wider surroundings. The Fourier series expansion is stated in terms of this wider concept of function. The coefficients of a Fourier series, as shown by the Euler-Lagrange result, are obtained by integration and not by differentiation as in the case of the Taylor series. Each Fourier coefficient A_n is obtained by integrating $f(x) \sin nx$ over the entire range. Thus any modification of $f(x)$ in a limited portion of the range changes all of the Fourier coefficients. It follows that the interconnections operate in the Fourier series in a global sense and not in a local sense as in the case of the Taylor series. It is the behavior of $f(x)$ *in the large* that matters in the case of the Fourier series, and not so much the behavior in the vicinity of a point. How can we resolve the differences between these two types of expansions: the Taylor series, which is the expansion about a point which gives strict predictions a finite distance from the point, and the Fourier series, which is an expansion in the large and which

gives knowledge of the function in the entire range. The Taylor series requires unlimited differentiability at a point, whereas the Fourier series does not demand any differentiability properties whatever.

Surprisingly enough, the chasm between the Taylor series and the Fourier series is bridged by means of the z-transform, which is the fundamental transform used in the theory of digital signal processing. Let us consider an analytic function $f(z)$ of the complex variable

$$z = x + jy = (re^{j\theta})^{-1}.$$

We now expand the function in a Taylor series (in the variable z^{-1}) about the point $z^{-1} = 0$, to obtain the z-transform

$$f(z) = \sum_{n=0}^{\infty} a_n z^{-n}.$$

The radius of convergence of this series extends from $z^{-1} = 0$ to the first singular point, say, z_0^{-1}. A singular point is a point where the function ceases to be analytic. The region of convergence of the Taylor series expansion of $f(z)$ is the region in the z-plane outside the circle of radius $|z_0|$; that is, the region of convergence is for all points z such that $|z^{-1}| < |z_0^{-1}|$ or equivalently $|z| > |z_0|$.

Let us now write the Taylor series expansion for points on the unit circle $z = \cos\theta - j\sin\theta$. We have

$$f(e^{-j\theta}) = \sum_{n=0}^{\infty} a_n(\cos n\theta - j\sin n\theta)$$

which is in the form of a complex Fourier series in the angle θ. Three cases can occur. In the first case, the singular point z_0 is inside the unit circle in the z-plane. In this case, the function is analytic on the unit circle and the Fourier series thus is an analytic representation of this analytic function. The French Academy believed this was the only case. In the second case, the singular point z_0 is outside the unit circle. In this case, the Taylor series does not represent the function, and so we will not consider the case further. The third case is the interesting one, and is the case which resolves the mathematical controversy which led up to Fourier's discovery in 1807. When the singular point z_0 lies on the unit circle, the Taylor series will not converge at some or all of the points on the unit circle. Thus the Taylor series defines an analytic function, which is differentiable to any order outside the unit circle, but the function becomes nonanalytic at some or all of the points on the unit circle. The Fourier series in θ is the Taylor series for z on the unit circle, and thus the Fourier series represents a function in the variable θ, which is nonanalytic at some or all of the points in its range $-\pi \leqslant \theta \leqslant \pi$. A small modification of the Fourier coefficients that would move the singular point z_0 from on the unit circle to just inside the unit circle would change a nonanalytic Fourier representation to an analytic one. The amazing thing is that it is enough to move the singularity from the periphery of the unit circle to the inside by an arbitrarily small amount, in order to change the given nondifferentiable function in θ to one which can be differentiated any number of times. Thus the mistake of the great French mathematicians of the prestigious French Academy who wanted to restrict the validity of Fourier series to analytic functions depended entirely on that extremely small but finite distance from a point on the periphery to a point just inside the unit circle. A function can be extremely smooth right up to the unit circle, and then disintegrate into a rough and distorted image of its former self once it is on the unit circle. The Taylor series breaks down on the unit circle, but its counterpart, the Fourier series in θ, is still valid. The theorem of Fourier is true; science could blossom.

V. The Sturm-Liouville Spectral Theory of Differential Equations

Following the great innovation of Fourier in 1807, the remarkable properties of Fourier series were gradually developed throughout the nineteenth century and into the twentieth century. The Fourier series as introduced by Fourier is an expansion in terms of cosines and sines, which represent an orthogonal set of functions. However, there are many other sets of orthogonal functions, and so today any such expansion in terms of orthogonal functions is called a Fourier series. As we will see, some sets of orthogonal functions can be stochastic, and it turns out that the corresponding Fourier series play an important role in statistical spectral analysis.

First, however, let us look at the important generalizations made by the French mathematicians Charles Sturm (1803–1855) [7] and Joseph Liouville (1809–1882) [8] in the decade of the 1830's. Let us now briefly look at the Sturm–Liouville theory of differential equations. The vibration of any infinitely long right circular cylinder of radius one can be described by a second-order differential equation. Let us consider a simple case, namely, the differential equation (the one-dimensional *Helmholtz equation*)

$$u''(x) + k^2 u(x) = 0.$$

The Helmholtz equation can be obtained by taking the temporal Fourier transform of the wave equation, which set off the search for the theory of Fourier. Here k is the wavenumber which is equal to ω/c where ω is the temporal frequency. In the Helmholtz equation, k^2 is some undetermined parameter. The variable x is the central angle of the cylinder, and so x lies in the range $-\pi$ and π. Because the points $x = -\pi$ and $x = \pi$ represent the same point on the cylinder, we must have the two boundary conditions

$$u(-\pi) = u(\pi)$$

$$u'(-\pi) = u'(\pi).$$

The general solution of the differential equation is

$$u(x) = A\cos kx + B\sin kx.$$

The two boundary conditions restrict the choice of the parameter k^2 to the discrete set of values

$$k^2 = 0, 1^2, 2^2, 3^2, \cdots$$

which are called the *eigenvalues* of the Helmholtz equation. The corresponding solutions of the equation, namely, the functions

$$u_k(x) = A\cos kx + B\sin kx$$

are called the *eigenfunctions*. These eigenfunctions are the cosine and sine functions which Fourier had used to construct his Fourier series. These functions represent the characteristic vibrational modes of the cylinder, which can only vibrate in this discrete set of wavenumbers $k = 0, 1, 2, \cdots$. Thus the Sturm–Liouville theory has given the answer to why the discrete set of cosine and sine functions were the correct ones for Fourier to use in a problem which stemmed from the wave equation.

Furthermore, the Sturm–Liouville theory gives us added insight to spectral analysis and, in fact, is the foundation of the spectral theory of differential equations. Most of the eigenvalue problems of mathematical physics are characterized by differential operators H of the form

$$H = \frac{d}{dx}\left[A(x)\frac{d}{dx}\right] + B(x).$$

The physical problems we consider require that the function $A(x)$ be positive within the given interval. Let us now form the operation $vHu - uHv$, which is

$$vHu - uHv = \frac{d}{dx}[A(x)(vu' - uv')].$$

We notice that the right-hand side is a total derivative, and so we have

$$\int_a^b (vHu - uHv)\,dx = [A(x)(vu' - uv')]_a^b.$$

Any differential operator H, which allows the transformation of such an integral (as on the left) into a pure boundary term (as on the right), is called *self-adjoint*. Thus the Sturm–Liouville operator H is self-adjoint. Often we may prescribe boundary conditions so that the right-hand side vanishes; such boundary conditions are called self-adjoint. We then have a self-adjoint problem, namely, a problem characterized by a self-adjoint operator H and self-adjoint boundary conditions. We then have the identity in the functions $u(x)$ and $v(x)$ given by

$$\int_a^b (vHu - uHv)\,dx = 0$$

which is called *Green's identity*.

The eigenvalue problem associated with the self-adjoint operator H starts with the differential equation

$$H\phi = \lambda\phi.$$

A solution satisfying the boundary conditions does not exist for all values of λ, but only for a certain selected set λ_i called the eigenvalues. This set consists of an infinite number of eigenvalues λ_i which are all real and which tend to infinity with i. We generally arrange these eigenvalues in increasing order to obtain the infinite sequence (called the *spectrum*)

$$\lambda_1, \lambda_2, \lambda_3, \cdots$$

together with the corresponding eigenfunctions

$$\phi_1, \phi_2, \phi_3, \cdots.$$

We now consider two different eigenvalues λ_j, λ_k and their corresponding eigenfunctions ϕ_j, ϕ_k. If we substitute $u = \phi_j$ and $v = \phi_k$ into Green's identity, we obtain

$$\int_a^b (\lambda_j \phi_j \phi_k - \lambda_k \phi_k \phi_j)\,dx = 0$$

which gives the orthogonality condition

$$\int_a^b \phi_j(x)\,\phi_k(x)\,dx = 0, \quad \text{for } j \neq k.$$

By normalization, we can require that

$$\int_a^b \phi_j^2(x)\,dx = 0$$

so that the eigenfunctions form an *orthonormal* set. The orthonormal property can be written more concisely as

$$\int_a^b \phi_j(x)\,\phi_k(x)\,dx = \delta_{jk}$$

where δ_{jk} is the Kronecker delta function.

Let us now represent an arbitrary function $f(x)$ in the form of the infinite expansion

$$f(x) = \sum_{k=1}^{\infty} c_k \phi_k(x).$$

As we have previously mentioned, such an expansion is called a *Fourier series* in honor of the pioneering work of Fourier. The Fourier coefficients c_k are obtained by multiplying both sides by $\phi_j(x)$ and integrating. The result is

$$c_j = \int_a^b f(x)\,\phi_j(x)\,dx.$$

Under certain general conditions, it can be shown that the orthonormal set is complete, so that the above Fourier expansion actually converges to the function $f(x)$. Suppose now that $f(x)$ is the solution to the inhomogeneous differential equation

$$Hf(x) = p(x).$$

In terms of linear system theory, $p(x)$ is the input and $f(x)$ is the output. Now substitute $u = f$ and $v = \phi_k$ into Green's identity. We obtain

$$\int_a^b (\phi_k Hf - fH\phi_k)\,dx = 0$$

which is

$$\int_a^b (\phi_k p - f\lambda_k \phi_k)\,dx = 0.$$

The above equation can be written as

$$\int_a^b f\phi_k\,dx = \frac{1}{\lambda_k}\int_a^b \phi_k p\,dx.$$

We recognize the left-hand side as the expression for the Fourier coefficient c_k. Thus

$$c_k = \frac{1}{\lambda_k}\int_a^b \phi_k(\xi)\,p(\xi)\,d\xi.$$

We now substitute this expression for c_k into the Fourier series to obtain

$$f(x) = \sum_{k=1}^{\infty} c_k \phi_k(x) = \int_a^b p(\xi)\left[\sum_{k=1}^{\infty}\frac{\phi_k(x)\,\phi_k(\xi)}{\lambda_k}\right]d\xi.$$

If we denote the expression in brackets by $G(x, \xi)$, then this equation is

$$f(x) = \int_a^b p(\xi)\, G(x, \xi)\, d\xi.$$

This is the integral form of the input–output relationship, and we recognize

$$G(x, \xi) = \sum_{k=1}^{\infty} \frac{\phi_k(x)\, \phi_k(\xi)}{\lambda_k}$$

as the *impulse response function* or *Green's function* (under the given boundary conditions), a concept originated by George Green (1793–1841) [9]. This equation exhibits the impulse response function of a linear system in terms of its spectrum $\lambda_1, \lambda_2, \lambda_3, \cdots$. We can confirm that the Green's function is indeed the impulse response by setting the input $p(x)$ equal to the impulse $\delta(x - x_0)$. Then the output is

$$\int_a^b \delta(\xi - x_0)\, G(x, \xi)\, d\xi = G(x, x_0)$$

and so $G(x, x_0)$ represents the output at x due to an impulse at x_0. Since the differential equation represents an input–output system, we see that the Green's function satisfies

$$HG(x, x_0) = \delta(x - x_0).$$

This equation shows that the Green's function $G(x, x_0)$ is the inverse of the differential operator H.

We have thus reviewed the spectral theory of differential operators, and now we can look at the most spectacular application of spectral estimation—quantum physics.

VI. Schrödinger Spectral Theory of the Atom

The Sturm–Liouville theory of the expansion of functions in terms of orthogonal functions found numerous physical applications in the work of Lord Rayleigh (1842–1919). Such expansions occur throughout the study of the elastic vibrations of solids and in the theory of sound. In the history of physics, a decisive breakthrough occurred when Erwin Schrödinger (1887–1961) [10] showed in 1926 that the vibrations occurring within the atom can be understood by means of the Sturm–Liouville theory. Let us now explain how the wave mechanics of Schrödinger describes the spectral lines of the atom. An equivalent matrix mechanics was formulated a year before Schrödinger by Werner Heisenberg (1901–1976) [11].

Before quantum theory, classical physics was at an impasse. It could not explain the existence of atomic spectra. For example, the bright yellow spectral line of sodium discovered by Bunsen means that the radiation of its atoms produces a discrete frequency ω_0. If we assume that this line is emitted by an electron, then the laws of classical physics state that such an electron should emit not a discrete line at ω_0, but a whole spectrum of lines at all frequencies ω, and with no discontinuities in the spectrum. That is, classical physics predicts that the spectrum of an electron should be continuous as is the spectrum of the sun. Yet Bunsen observed the discrete spectrum of sodium as evidenced by the bright yellow line. (As we will soon see, this line observed by Bunsen is actually a doublet, which Bunsen was unable to resolve with the means available to him.)

Quantum mechanics allows us to see the atom from a new point of view. Quantum mechanics says that atomic electrons jump from one energy state to another, and that the difference of these energies is embodied as a quantum of electromagnetic energy, the photon. If the energy diminishes, a photon is born. If the energy increases, a photon or a quantum of energy from some other field has been absorbed just before the jump.

In quantum mechanics, an electron is represented by a probability density function. (The probability density function is found as the squared magnitude $|\phi|^2$ of a probability wave function ϕ.) An electron jump has a probability that depends upon the shapes of the probability density functions that correspond to the electron prior to and after the jump. The probability of a jump is, generally speaking, greater for the stronger overlapping or deeper interpenetration of these probability density functions. The laws that divide electron transitions in atoms into more probable and less probable ones are called selection rules. It is in this jumping of electrons that photons are born. These photons enter a spectroscope, get sorted into types, and produce the spectral lines.

The more photons that an atom emits in a second, the brighter the spectral lines. If the number of atoms remains constant, then the brightness of the spectral lines depends upon the statistical frequency of electron jumps in the atoms. And this statistical frequency is determined by the probability distribution of jumps. It is in this way that an atomic spectrum consisting of a number of lines of different brightnesses is generated.

One can make the observation that the spectrum estimation problem (the subject matter of this special issue of *Proceedings of the IEEE*) is not central to the spectral representation in quantum mechanics. This situation was brought forcibly to the writer's attention several years ago at the U.S. Air Force Geophysics Library at Hanscom Field, MA, which is one of the best scientific libraries in the world. The many shelves devoted to "spectra" consisted of a mixture of both kinds of books, but no book devoted to a discussion of the relationship between the two areas of spectral theory.

Spectral estimation in quantum mechanics is based on the edifice of spectroscopy, which is an instrumentational science. In 1891, the physicist A. A. Michelson developed an interferometer, a device producing the superposition of a light signal on top of itself with a prescribed delay. In one series of experiments, Michelson first bandpass filtered a light signal by passing it through a prism. He then used the interferometer to measure the visibility of the superimposed signal as a function of delay. The resulting curve was the autocovariance function of the original signal. Michelson then used a mechanical harmonic analyzer to compute the Fourier transform of the visibility curve; that is, he estimated the power spectrum of the signal. Michelson's experiments were done to examine the fine structure of spectral lines of light. Thus in those early days, the present day dichotomy of spectrum estimation had not yet materialized.

The technique of spectral analysis in physics developed rapidly in the twentieth century, and the instruments became more powerful and sensitive. The spectroscopists came up with the following question for theoreticians, namely, the question of why spectral lines are somewhat fat, not infinitesimally thin.

It was recognized that a photon corresponds to a line at one frequency ω. The question was why the lines on a photographic plate of a spectroscope come out somewhat broadened, not slender. The answer was found in the wave property of the electron and the Heisenberg uncertainty principle. The initial energy of an electron in an atom refers to a

stationary state, and so does the final energy. However, an electron jump is in violation of some steady state. As soon as this occurs, the Heisenberg principle takes over. If we let Δt designate the lifetime of an electron between jumps, then the uncertainty of photon energy is $\Delta E \sim h/\Delta t$, where h is Planck's constant. Using Planck's formula for energy quanta, the uncertainty ΔE of the energy is proportional to the uncertainty $\Delta\omega$ of the frequency of the photon

$$\Delta E = \frac{h}{2\pi}\Delta\omega.$$

Thus the spectral lines have a width $\Delta\omega$ which is inversely proportional to the time of the "settled life" of the electron in the atom

$$\Delta\omega \sim \frac{2\pi}{\Delta t}.$$

In other words, the more "settled" or quiescent the life of the electron in the atom, the narrower the spectral lines. That is why at high temperatures and pressures, when many of the atomic electrons are unsettled, the spectral lines broaden out and become smeared. Thus an individual spectral line has a finite width associated with thermal motion and collision broadening. This is not only important in physics, but it relates very importantly to the topic of spectrum estimation in this special issue. Real "lines" have finite width. This means that real lines behave like narrow-band noises and not like either single frequencies or a constant-amplitude lightly frequency-modulated signal.

Let us now return the discussion of the yellow sodium line which Bunsen observed. The sodium D line is a doublet. Moreover the sodium spectrum contains four lines in the visible range, and two more in the near ultraviolet, strong enough to be useful for analytic chemistry. The sodium spectrum contains 29 lines of astrophysical interest between the D lines and 4390 Å (still in the visible).

We might say that Bunsen over a century ago was performing spectrum estimation. He was unable to resolve the two frequencies present in the doublet, even as today a person doing spectrum estimation might have the same problem in some other situation. Also Bunsen missed the many other lines in the sodium atom, even as today a person doing spectrum estimation might not find some features without the use of modern techniques. As spectrosopic instruments became better, these lines were discovered. Now another question, however, has come up. Many spectral lines, which, it would seem, should correspond to a single frequency, actually turned out to be the states of a number of very close-lying lines. The fact that the sodium D line is a doublet is a case in point. The fine structures of spectral lines (doublets, etc.) were revealed only because of the great advances in spectral techniques. In turn, electron spin was discovered in order to explain these "fine qualities" in spectra. Let us briefly give the reason. When spectra are generated, the states of two electrons with opposite spins can have slightly different energies. As a result, the spectral line is doubled; in place of one line we have twin lines with identical brightnesses. Such twins are usually born only when the outer electron shell has one electron. If the number of electrons in this shell increases, we can have triplets and even larger families of the former spectral line.

Let us now consider the quantum mechanical formulation of the harmonic oscillator problem. In terms of the nondimensional displacement x, the time-independent *Schrödinger equation* is

$$H\phi = \lambda\phi$$

where H is defined as the differential operator

$$H = \frac{d^2}{dx^2} - x^2$$

and λ is defined as

$$\lambda = \frac{2E}{\hbar\omega_0}.$$

Here ϕ is the probability wave function, the constant E is the energy, $h = 2\pi\hbar$ is Planck's constant, and the constant ω_0 is the natural frequency. The problem of finding the probability wave function ϕ is a Sturm–Liouville problem. The solution gives the eigenvalues as 1, 3, 5, 7, $\cdots$, and so we write

$$\lambda_k = (2k+1), \quad \text{for} \quad k = 0, 1, 2, \cdots.$$

Thus the eigenenergies are

$$E_k = \tfrac{1}{2}\hbar\omega_0\lambda_k = \hbar\omega_0(k+\tfrac{1}{2}), \quad \text{for} \quad k = 0, 1, 2, \cdots.$$

The corresponding eigenfunctions are

$$\phi_k = C_k h_k(x)\, e^{-x^2/2}, \quad \text{for} \quad k = 0, 1, 2, \cdots$$

where C_k is a normalization constant, and $h_k(x)$ is the Hermite polynomial of order k. The discrete set of eigenenergies E_0, $E_1, E_2, \cdots$ represent the discrete lines observed in the spectrum. Thus quantum mechanics, through the use of Sturm–Liouville theory, is able to explain the existence of atomic spectra. However, certain mathematical difficulties remained; the history of their resolution is given in the next section.

VII. The von Neumann Spectral Representation Theorem

In finite-dimensional space, the following eigenvalue problem is posed. Given an Hermitian matrix H, find all column-vector solutions ϕ of the characteristic equation

$$H\phi = \lambda\phi$$

where λ is a constant also to be determined. That is, given H, find ϕ and λ. The solutions $\phi_1, \cdots, \phi_n$ are called the eigensolutions (assumed to be normalized), and the corresponding real numbers $\lambda_1, \cdots, \lambda_n$ are called the eigenvalues of the matrix H. The totality of the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, in order of increasing magnitude, is called the spectrum. Now write the eigenequations

$$H\phi_k = \lambda_k\phi_k \quad (\text{for} \quad k = 1, \cdots, n)$$

in the form of the matrix equation

$$HU = U\Lambda.$$

Because the eigensolutions are orthonormal, the matrix U (which has the eigensolutions as it columns) is unitary, i.e.,

$$UU^T = I$$

where I is the identity matrix. (The superscript T indicates complex conjugate transpose.) The matrix Λ is diagonal matrix, with the spectrum along its diagonal. Thus this eigenvalue problem can be described as the problem of finding a unitary matrix U that reduces H to a real diagonal matrix,

i.e.,

$$U^{-1}HU = \Lambda.$$

(Note: In case H is real, then H is a symmetric matrix and U is an orthogonal matrix.)

Although the unitary matrix U, whose columns are the eigensolutions ϕ_i, is not uniquely determined by H, John von Neumann [12] in 1929 exploited the unitary nature of U to reformulate the eigenvalue problem. The von Neumann reformulation, which is called the *spectral representation problem*, yields the same results as the eigenvalue problem in finite-dimensional space, but has the advantage that it can be extended to Hilbert space.

We recall that the diagonal matrix Λ is defined to be the matrix with the eigenvalues, ordered by increasing magnitude, along its main diagonal and zeros off the diagonal. Because of this ordering, the matrix Λ is uniquely determined for any given Hermitian matrix H. Because some eigenvalues may be repeated, let us relabel them as $\lambda_1, \lambda_2, \cdots, \lambda_m$ (with $m \leqslant n$), where each λ_i is now distinct. Consequently for a given H, we have the unique decomposition

$$\Lambda = \lambda_1 Q_1 + \lambda_2 Q_2 + \cdots + \lambda_m Q_m$$

where Q_i is a diagonal matrix with 1's in those places on its main diagonal in which λ_i occurs in Λ and 0's elsewhere. The sum of the Q_i gives the identity matrix

$$I = Q_1 + Q_2 + \cdots + Q_m.$$

We now define the matrix P_j as

$$P_j = UQ_jU^{-1} \qquad (\text{for } j = 1, 2, \cdots, m).$$

A projection matrix is defined as a Hermitian idempotent matrix. Because Q_j is Hermitian ($Q_j = Q_j^T$) and idempotent ($Q_jQ_j = Q_j$), it follows that Q_j is a projection matrix. Because P_j is Hermitian ($P_j = P_j^T$) and idempotent

$$P_jP_j = UQ_jU^{-1}UQ_jU^{-1} = UQ_jQ_jU^{-1} = P_j$$

it follows that P_j is a projection matrix. Since for $i \neq j$

$$P_iP_j = UQ_iU^{-1}UQ_jU^{-1} = 0$$

it follows that $P_i + P_j$ is a projection matrix and the space spanned by P_i is orthogonal to the space spanned by P_j. Let us now define the function $\mathcal{H}(\lambda)$ of the continuous variable λ as

$$\mathcal{H}(\lambda) = P_1\delta(\lambda - \lambda_1) + P_2\delta(\lambda - \lambda_2) + \cdots + P_m\delta(\lambda - \lambda_m).$$

This function is the continuous representation of the suite of projection matrices $P_1, P_2, \cdots, P_m$.

We now consider the quadratic form uHv where u is a row vector and v is a column vector. We have

$$\begin{aligned} uHv &= uU\Lambda U^{-1}v = uU(\lambda_1 Q_1 + \lambda_2 Q_2 + \cdots + \lambda_m Q_m)U^{-1}v \\ &= u(\lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_m P_m)v \\ &= \lambda_1 uP_1v + \lambda_2 uP_2v + \cdots + \lambda_m uP_mv. \end{aligned}$$

The essence of the von Neumann spectral representation lies in the fact that the components uP_jv are numerically invariant for given u, H, and v. In this way, the nonuniqueness of the unitary matrix U appearing in the eigenvalue decomposition is bypassed. We see that we can write the quadratic form as the integral

$$uHv = \int_{-\infty}^{\infty} \lambda u\mathcal{H}(\lambda)v\, d\lambda.$$

This equation represents the von Neumann *spectral representation of the Hermitian matrix H.*

Let us now analyze this equation. If we strip the u and v from this equation, we are left with

$$H = \int_{-\infty}^{\infty} \lambda\mathcal{H}(\lambda)\, d\lambda$$

which, in matrix notation, is

$$H = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_m P_m.$$

We can write the row vector u as

$$u = \int_{-\infty}^{\infty} u\mathcal{H}(\lambda)\, d\lambda$$

which is

$$u = uP_1 + uP_2 + \cdots + uP_m.$$

Finally, we can write

$$Hv = \int_{-\infty}^{\infty} \lambda\mathcal{H}(\lambda)v\, d\lambda$$

which is

$$Hv = \lambda_1 P_1 v + \lambda_2 P_2 v + \cdots + \lambda_m P_m v.$$

Let us now consider functions of the matrix H. First, we consider the square of H. We have

$$\begin{aligned} H^2 &= (\lambda_1 P_1 + \cdots + \lambda_m P_m)^2 = \lambda_1^2 P_1 + \cdots + \lambda_m^2 P_m \\ &= \int_{-\infty}^{\infty} \lambda^2\mathcal{H}(\lambda)\, d\lambda. \end{aligned}$$

We see that squaring H results in squaring the λ inside the integral. In general, if we form a function of H, then the result is that the same function of λ is taken within the integral sign; that is

$$f(H) = \int_{-\infty}^{\infty} f(\lambda)\,\mathcal{H}(\lambda)\, d\lambda.$$

The above spectral representation was derived for finite-dimensional space, that is, a space in which the elements u and v are vectors and the Hermitian operator H is a matrix. One of the major achievements of von Neumann was the development of the concept of the infinitely-dimensional space, which he called *Hilbert space* in honor of the great mathematician David Hilbert (1862–1943). We now let u and v represents elements in Hilbert space, and let H represent a Hermitian operator. A Hilbert space is characterized by an inner product (or dot product). The *inner product* of the elements u and v is denoted by $\langle u, v\rangle$. If we let H operate on the element v, we obtain a new element Hv. The inner product of the elements u and Hv is denoted by $\langle u, Hv\rangle$. This inner product is the counterpart of the quadratic form uHv in finite-dimensional space. Once we establish this connection, it turns out that the von Neumann spectral representation has exactly the same form in Hilbert space as it does in finite-dimensional space. Thus in Hilbert space, we also have an operator $\mathcal{H}(\lambda)$, which is the continuous representation of the suite of projection operators associated with the Hermitian operator H. Whereas in finite-dimensional space, we made use of the quadratic form $u\mathcal{H}(\lambda)v$, we now make use of its counterpart $\langle u, \mathcal{H}(\lambda)v\rangle$ in Hilbert space. Thus the von

Neumann *spectral representation* in Hilbert space is

$$\langle u, Hv\rangle = \int_{-\infty}^{\infty} \lambda \langle u, \mathcal{H}(\lambda)\, v\rangle\, d\lambda.$$

Let us now look at some history. In general, there is no quadratically integrable solution to the eigenvalue problem in Hilbert space. This circumstance, however, bothered no one working in physics. Wavelet solutions (i.e., quadratically integrable superpositions of eigenfunctions with eigenvalues in a small neighborhood) were used from the start, appearing in the works of de Broglie and Schrödinger from 1924.

One of the authors cited in the Reference Section knew von Neumann personally, studied his work assiduously, and certainly regards him as one of the truly great founders of quantum theory. However, there was never a "crisis in physics" that was resolved by the von Neumann spectral representation theorem. Most people doing the practical calculations to be compared with experiment had never heard of the theorem, which was for them at such a high level of abstraction that it had no bearing on what they were doing.

Throughout this essay we have traced the development of spectral theory, from the analytic functions of Brook Taylor, to the nondifferentiable functions of Jean Baptiste Joseph de Fourier, and now to the more general operators of Hilbert space. At each stage, these developments were mathematical in nature, but they laid the foundations for subsequent advances in physics. Reasoning in mathematics and reasoning in physics often appear quite different. When a major physical breakthrough occurs, such as in quantum mechanics in the 1920's, and a flood of exciting new physical results come out, certainly the work of mathematicians in establishing existence and uniqueness theorems might seem somewhat irrelevant.

For a moment let us go back to Sir Isaac Newton. It is often said that the unique greatness of Newton's mind and work consists in the combination of a supreme experimental with a supreme mathematical genius. It is also often said that the distinctive feature of Newtonian science consists precisely in the linking together of mathematics and experiment, that is, in the mathematical treatment of experimental or (as in astronomy, geophysics, or wherever experiments cannot be performed) observational data. Yet, although correct, this description does not seem to be quite complete; there is more in the work of Newton than mathematics and experiment. There is also a deep intuition and insight in his interpretation of nature.

In today's science, specialization has gone far. Physicists use mathematics; they formulate problems, devise methods of solution, and perform long derivations and calculations, but generally they are not interested in creating new mathematics. The discovery and purification of abstract concepts and principles is particularly in the realm of mathematics. John von Neumann (1903–1957) is a prime example of a mathematician doing physics. When he did physics, he thought and calculated like a physicist, only faster. He understood all branches of physics, as well as chemistry and astronomy, but mainly he had a talent for introducing only those mathematical ideas that were relevant to the physics at hand. The introduction of abstract Hilbert space theory in quantum mechanics, chiefly by von Neumann, made possible the construction of a solid theory on the basis of the powerful intuitive ideas of Dirac and other physicists.

The physics of quantum theory cannot be mathematically formulated in finite-dimensional space but requires Hilbert space. After the work of Heisenberg and Schrödinger in 1925 and 1926, there was a crisis in abstract mathematics because the physics of quantum mechanics could not be adequately formulated in terms of the existing mathematical framework. This situation was rectified in 1929 by von Neumann [12] who laid the mathematical foundations of quantum mechanics in terms of Hilbert space. There is an apocryphal story that the young John von Neumann, who was barely past being a teenager, and had not yet earned his doctorate, was lecturing in Göttingen. Of course, most of the famous physicsts present regarded his work as too abstract, but the great mathematician Hilbert was in the audience. As the story goes, the elderly Hilbert leaned over and whispered into Professor Courant's ear: "What is this Hilbert space?" Another even more apocryphal story goes as follows. A group of physicists came to von Neumann and described a problem in physics which they could not solve. After thinking for a while, von Neumann in his head came up with the numerical answer which agreed with the experimental result, which the physicists knew but had not told him. They were very impressed and they blurted out "Dr. von Neumann, the general solution involves solving an infinite set of nonlinear partial differential equations. Certainly you have found some mathematical shortcut!" von Neumann answered "No, I solved the infinite set."

von Neumann [13] showed that from a mathematical point of view, it is the spectral representation that is required in quantum mechanics rather than the solution of the eigenvalue problem as such. In this sense, spectral theory represents the key to the understanding of the atom. In fact, von Neumann [13] has shown that the spectral representation enters so essentially into all quantum mechanical concepts that its existence cannot be dispensed with. His establishment of the spectral representation of the Hermitian operator H is one of the great achievements in mathematics, and a milestone in the history of spectral theory.

VIII. Einstein-Wiener Theory of Brownian Motion

A highly interesting kinetic phenomenon known as Brownian movement was first reported in 1827 by the distinguished botanist, Robert Brown, who found that "extremely minute particles of solid matter when suspended in pure water exhibit motions for which I am unable to account and which, from their irregularity and seeming independence, resemble in a remarkable degree, the less rapid motions of some of the simplest animalcules of infusions." This type of irregular zigzag movement is typified by the dancing of dust particles in a beam of light. The cause of Brownian movement was long in doubt, but with the development of the kinetic theory of matter came the realization that the particles move because they are bombarded unequally on different sides by the rapidly moving molecules of the fluid in which they are suspended. The Brownian movement never ceases. The detailed physical theory of Brownian movement was worked out in 1904 by M. von Smoluchowski [14], and in a more final form in 1905 by Albert Einstein [15]. In 1923, Norbert Wiener [16] developed the mathematical theory of Brownian movement, which today is the basis of the mathematical model of white noise in continuous time. White noise is defined as a stationary random process which has a constant spectral power density. The concept of the white noise process, as given by the Einstein-Wiener theory of Brownian motion, is important in all theoretical studies of spectrum analysis.

In practice, a signal is of finite duration, and usually can be digitized on a grid fine enough for interpolation to be adequate. In this sense, the set of data representing a signal is really finite. Accordingly, we do not have to go to continuous

time or to infinite time unless 1) we so wish or 2) we gain from it. In other words, as long as we stay finite, we do not need the Einstein-Weiner theory. With this *caveat emptor*, let us now discuss this theory.

A white noise process in continuous time cannot be represented by the ordinary types of mathematical functions which one meets in calculus. Instead, white noise can only be represented by what mathematicians call a *generalized function*. The most familiar example of generalized function is the *Dirac delta function*, which is often defined as

$$\delta(t - t_0) = \begin{cases} 0, & \text{for } t \neq t_0 \\ \infty, & \text{for } t = t_0 \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(t - t_0)\, dt = 1.$$

The most important property of the delta function is its sifting property, that is, its ability to isolate or reproduce a particular value of an ordinary function $f(t)$ according to the convolution formula

$$\int_{-\infty}^{\infty} f(t - t_0)\, \delta(t)\, dt = f(t_0).$$

If one feels uncomfortable with generalized functions, then one can often avoid them by using Lebesgue-Stieltjes integrals. For example, the Heaviside step function $H(t)$ is an ordinary function equal to zero for $t < 0$ and to one for $t \geqslant 0$. Since

$$dH(t) = \delta(t)\, dt$$

the above convolution formula becomes the Lebesgue-Stieltjes integral

$$\int_{-\infty}^{\infty} f(t - t_0)\, dH(t) = f(t_0).$$

This Lebesgue-Stieltjes integral involves only ordinary functions.

Let us now look at a *white noise process* which we denote by $\epsilon(t)$. It is a generalized random function. Again let $f(t)$ be an ordinary function, and consider the convolution integral

$$\int_{-\infty}^{\infty} f(t - t_0)\, \epsilon(t)\, dt.$$

Let $\xi(t)$ be the integrated white noise process, so that we may write

$$d\xi(t) = \epsilon(t)\, dt.$$

The integrated white noise process $\xi(t)$ is an ordinary random function, and the above convolution becomes the Legesgue-Stieltjes integral

$$\int_{-\infty}^{\infty} f(t - t_0)\, d\xi(t).$$

Wiener formulated everything in terms of Lebesgue-Stieltjes integrals with ordinary functions. However, we are going to take a strictly engineering approach and formulate things in terms of ordinary integrals, but with generalized functions.

Without loss of generality in the discussion which follows, we can for convenience let $t_0 = 0$, so that the integral in question becomes

$$\int_{-\infty}^{\infty} f(t)\, \epsilon(t)\, dt.$$

As is usual statistical practice, let $\boldsymbol{E}$ denote the mathematical expectation operator. Since this operator is linear, it may be interchanged with integral signs (provided certain regularity conditions hold). The expectation of the above integral is

$$\boldsymbol{E} \int_{-\infty}^{\infty} f(t)\, \epsilon(t)\, dt = \int_{-\infty}^{\infty} f(t)\, \boldsymbol{E}\epsilon(t)\, dt.$$

Because we want white noise to have zero mean, we let $\boldsymbol{E}\epsilon(t) = 0$, and so the above integral is zero. Let us next consider the variance given by

$$\boldsymbol{E}\left[\int_{-\infty}^{\infty} f(t)\, \epsilon(t)\, dt\right]^2 = \boldsymbol{E}\left[\int_{-\infty}^{\infty} f(t)\, \epsilon(t)\, dt \int_{-\infty}^{\infty} f(\tau)\, \epsilon(\tau)\, d\tau\right]$$

$$= \int_{-\infty}^{\infty} f(t)\, f(\tau)\, \boldsymbol{E}[\epsilon(\tau)\, \epsilon(\tau)]\, dt\, d\tau.$$

Now we come to the key point. We want white noise to be uncorrelated at two different time points, but at the same time we want the variance of white noise to produce an impulse so as to make the above integral have a nonzero value. Thus the key element is to define the covariance $\boldsymbol{E}[\epsilon(t)\, \epsilon(\tau)]$ as being equal to $\delta(t - \tau)$. Then the above integral becomes

$$\int_{-\infty}^{\infty} f(t)\, f(\tau)\, \delta(t - \tau)\, dt\, d\tau = \int_{-\infty}^{\infty} f^2(t)\, dt.$$

We can therefore make the following definition. A generalized random function $\epsilon(t)$ is white noise provided that $\boldsymbol{E}\epsilon(t) = 0$ and $\boldsymbol{E}\epsilon(t)\, \epsilon(\tau) = \delta(t - \tau)$. For a long time such a random process was regarded as improper. As we know, the delta function can be approximated arbitrarily close by ordinary functions. Likewise, the white noise process $\epsilon(t)$ can be approximated arbitrarily close by ordinary random processes.

Because one never uses the white noise process in isolation but only in integrals, the white noise process can be avoided by the use of the Lebesgue-Stieltjes integral, just as the Dirac delta function can be so avoided. However, as we have said, we will not follow the Lebesgue-Stieltjes approach here.

Let us now consider white noise $\epsilon(n)$ for discrete (integer) time n. White noise in discrete time is not a generalized random process, for $\epsilon(n)$ is merely a sequence of zero-mean, constant-variance, uncorrelated random variables. However, the Fourier transform of discrete white noise is a generalized random process, which we denote by $E(\omega)$. We have

$$E(\omega) = \sum_{n=-\infty}^{\infty} \epsilon(n)\, e^{-j\omega n} \qquad (\text{for } -\pi \leqslant \omega \leqslant \pi).$$

We can easily verify that $E(\omega)$ has zero mean. The covariance of $E(\omega)$ is

$$\boldsymbol{E}[E^*(\omega)\, E(\mu)] = \boldsymbol{E} \sum_n \epsilon(n)\, e^{j\omega n} \sum_k \epsilon(k)\, e^{-j\mu k}$$

$$= \sum_n \sum_k [\boldsymbol{E}\epsilon(n)\, \epsilon(k)]\, e^{j(\omega n - \mu k)}.$$

Because $\boldsymbol{E}\epsilon(n)\, \epsilon(k) = \delta_{nk}$ (i.e., the Kronecker delta function,

which is one when $n = k$ and zero otherwise), we have

$$E[E^*(\omega)\,E(\mu)] = \sum_n e^{-jn(\mu-\omega)} = 2\pi\delta(\mu - \omega).$$

That is, the covariance is a Dirac delta function. Thus we come to an important result: The Fourier transform of a white noise process in (infinitely extended) discrete time n is a white noise process in the continuous variable ω. (It is easy to show that the corresponding result holds for the case of a white noise process in continuous time.) In other words, the Fourier transform of a very rough (white) process in time is a very rough (white) process in frequency. The Fourier transform preserves (saves) information and does not smooth (destroy) information. Today this result is second-nature to an engineer, but when Wiener obtained this result in 1923 it was startling. Wiener unlocked the spectral theory of the most random of processes (white noise), and now the stage was set for applying this result to the more smooth processes which are generated by many physical phenomena. Wiener made this application in 1930 under the name of generalized harmonic analysis, but before we give its history we will break our train of thought and look at the innovative work of Yule in 1927. Yule's work at the time seemed modest. While most mathematicians and physicists were developing general methods to deal with the infinite and the infinitesimal in spectrum analysis, Yule was developing a simple model with a finite number of parameters (i.e., a finite parameter model) in order to handle spectrum analysis in those cases where this model was appropriate. This model of Yule is known as the autoregressive (AR) process.

IX. Yule Autoregressive Spectrum Estimation Method

At the turn of the twentieth century, Sir Arthur Schuster [17] introduced a numerical method of spectrum analysis for empirical time series. Let $x(n)$ represent the value of a time series at discrete (integer) time n. Given N observations of the time series from $n = 1$ to $n = N$, then Schuster's method consisted of computing the *periodogram* $P(\omega)$ defined as

$$P(\omega) = \frac{1}{N}\,|x(1)\,e^{-j\omega} + x(2)\,e^{-j\omega 2} + \cdots + x(N)\,e^{-j\omega N}|^2.$$

For example, suppose that the time series consists of a sinusoid of frequency ω_0 with superposed errors; then, the periodogram would show a peak at $\omega = \omega_0$. Thus by computing the periodogram, the peaks would show the location of the frequencies of the underlying sinusoidal motion. Until the work of Yule (1871–1951) in 1927 [18], the Schuster periodogram approach was the only numerical method of empirical spectrum analysis. However, many empirical time series observed in nature yielded a periodogram that was very erratic and did not exhibit any dominant peaks. This led Yule to devise his autoregressive method of spectrum analysis. In those days, empirical spectrum analysis was called the investigation of periodicities in disturbed series. His main application was the determination of the spectrum of Wolfer's sunspot time series.

G. Udny Yule in 1927 introduced the concept of a finite parameter model for a stationary random process in his fundamental paper on the investigation of the periodicities in time series with special reference to Wolfer's sunspot numbers. If we consider a curve representing a sinusoidal function of time and superpose on the ordinate small random errors, then the only effect is to make the graph somewhat irregular, leaving the suggestion of periodicity still quite clear to the eye. If the errors are increased in magnitude, the graph becomes more irregular, the suggestion of periodicity more obscure, and we have only sufficiently to increase the errors to mask completely any appearance of periodicity. But, however large the errors, Schuster's periodogram analysis is applicable to such a time series, and given a sufficient number of observations should yield a close approximation to the period and amplitude of the underlying sinusoidal wave.

Yule reasoned in the following way. Consider a case in which periodogram analysis is applied to a time series generated by some physical phenomenon in the expectation of eliciting one or more true periodicities. Then it seemed to Yule that in such a case there would be a tendency to start with the initial hypothesis that the true periodicities are masked solely by additive random noise. As we well know, additive random noise does not in any way disturb the steady course of the underlying sinusoidal function or functions. It is true that the periodogram itself will indicate the truth or otherwise of the hypothesis made, but Yule saw no reason for assuming it to be the hypothesis most likely *a priori*.

At this point, Yule introduced the concept of an input–output feed-back model. The amplitude of a simple harmonic pendulum with damping (in discrete approximation) can be represented by the homogeneous difference equation

$$b(n) + a_1 b(n-1) + a_2 b(n-2) = 0.$$

Here $b(n)$ is the amplitude at discrete (integer) time n. Errors of observation would cause superposed fluctuations on $b(n)$, but Yule observed that, by improvement of apparatus and automatic methods of recording, errors of observation can be practically eliminated: An initial impulse or disturbance would set the pendulum in motion, and the solution of the difference equation would give the impulse response. The initial conditions are $b(n) = 0$ for $n < 0$ and $b(0) = 1$. The characteristic equation of the difference equation is

$$E^2 + a_1 E + a_2 = 0.$$

From physical considerations, we know that the impulse response is a damped oscillation, so the roots E_1 and E_2 of the characteristic equation must be complex with magnitude less than one. This condition is equivalent to the condition that $a_2 < 1$ and $4a_2 - a_1^2 > 0$. The solution of the difference equation thus comes out to be

$$b(n) = e^{\lambda n}\,\frac{\sin(n+1)\,\omega_0}{\sin\omega_0}$$

where

$$\lambda = 0.5 \ln a_2$$

$$\omega_0 = \tan^{-1}[-a_1^{-1}\sqrt{4a_2 - a_1^2}].$$

The damped oscillation $b(n)$ is the impulse response function. The frequency ω_0 is the fundamental frequency of the impulse response function.

As we mentioned above, Yule ruled out superposed errors. Now, however, he allows a driving function (or input) of white noise, which he describes in the following way. The apparatus is left to itself, and unfortunately boys get into the room and start pelting the pendulum with peas, sometimes from one side and sometimes from the other. He states that the motion is now affected, not by superposed fluctuations, but by driving disturbances. As a result, the graph will be of an entirely different kind than a graph in the case of a sinusoid with super-

posed errors. The pendulum and pea graph will remain surprisingly smooth, but amplitude and phase will vary continuously, as governed by the inhomogeneous difference equation

$$x(n) + a_1 x(n-1) + a_2 x(n-2) = \epsilon(n)$$

where $\epsilon(n)$ is the white noise input. The solution of this difference equation is

$$x(n) = \sum_{k=0}^{\infty} b(k)\, \epsilon(n-k)$$

where $b(k)$ is the impulse response function given above.

Yule thus created a model with a finite number of parameters, namely, the coefficients a_1 and a_2 of the difference equation. Given an empirical time series $x(n)$, he uses the method of regression analysis to find these two coefficients. Because he regresses $x(n)$ on its own past instead of on other variables, it is a self-regression or *autoregression*. The least squares normal equations involve the empirical autocorrelation coefficients of the time series, and today these equations are called the *Yule-Walker equations*.

Yule carried out his autoregressive analysis on Wolfer's sunspot numbers, which are a sequence of yearly observations of sunspot observations. He used the numbers over the period 1749–1924 and obtained the autoregressive equation (with the mean value removed)

$$x(n) - 1.34254\, x(n-1) + 0.65504\, x(n-2) = \epsilon(n)$$

and thus

$$\lambda = 0.5 \ln 0.65504 = -0.21154$$

$$\omega_0 = 33.963^\circ \text{ per year.}$$

Hence, the dominant period is $360^\circ/\omega_0 = 10.60$ years. Yule states that his autoregressive method represents an alternative method of estimating the spectrum, as opposed to the Schuster periodogram. In fact, his autoregressive model gives him an estimate not only of the power spectrum but of an amplitude-and-phase spectrum

$$B(\omega) = \sum_{n=0}^{\infty} b(n)\, e^{-j\omega n} = \frac{1}{1 + a_1 e^{-j\omega} + a_2 e^{-2j\omega}}$$

which, for the sunspot numbers, is

$$B(\omega) = \frac{1}{1 - 1.34254\, e^{-j\omega} + 0.65504\, e^{-2j\omega}}.$$

The magnitude $|B(\omega)|$ and the phase $\theta(\omega)$ are given by the equation

$$B(\omega) = |B(\omega)| e^{j\theta(\omega)}.$$

The power spectrum is the square of the magnitude spectrum, that is, $|B(\omega)|^2$. The peak is close to the fundamental frequency $\omega_0 = 33.963^\circ$ per year. Except in exploration geophysics [19], [20], where Yule's amplitude-and-phase spectrum $B(\omega)$ is physically the spectrum of the minimum-delay seismic wavelet, Yule's spectral estimation method received scant attention until the 1960's.

X. Wiener's Generalized Harmonic Analysis

Norbert Wiener [21] published in 1930 his classic paper, "Generalized Harmonic Analysis," which he personally considered his finest work. In his introduction, he states that he was motivated by the work of researchers in optics, especially that of Rayleigh and Schuster. However, Wiener demonstrated that the domain of generalized harmonic analysis was much broader than optics. Among Wiener's results were the writing down of the precise definitions of and the relationship between the autocovariance function and the power spectrum. The theorem that these two functions make up a Fourier transform pair is today known as the Wiener–Khintchine theorem [22].

Mention should be made of the basic fact that the existence of the spectrum follows from the properties of positive definite functions. Bochner's theorem on the spectral representation of positive definite functions provides a direct mathematical unification of spectral theories in Hilbert space and in stationary time series.

The writer several times in the 1950's discussed with Professor Wiener why his 1930 paper was not more accepted and used by the mathematical profession at the time. As with all things, Wiener looked at history quite objectively and with his characteristic concern for and love of people. In retrospect, it seems it was not until the publication of Wiener's book *Cybernetics* [23] in 1948 and also the nonclassified publication of his book *Time Series* [24] in 1949 that the general scientific community was able to grasp the overall plan and implications of Wiener's contributions.

The following passage from Wiener's 1933 book *The Fourier Integral* [25] indicates the philosophy of Wiener's thinking and his great personal appeal:

> "Physically speaking, this is the total energy of that portion of the oscillation lying within the interval in question. As this determines the energy-distribution of the spectrum, we may briefly call it the "spectrum." The author sees no compelling reason to avoid a physical terminology in pure mathematics when a mathematical concept corresponds closely to a concept already familiar in physics. When a new term is to be invented to describe an idea new to the pure mathematician, it is by all means better to avoid needless duplication, and to choose the designation already current. The "spectrum" of this book merely amounts to rendering precise the notion familiar to the physicist, and may as well be known by the same name."

Let us now define a stationary random process. We could either use discrete or continuous time, but for convenience let us use discrete (integer) time n. Let the process be denoted by $x(n)$, which, we will assume, has zero mean. The process is called (second-order) stationary, provided that its autocovariance function

$$\phi(k) = E x^*(n)\, x(n+k)$$

depends only upon the time-shift k. Here, as always, the superscript asterisk indicates complex-conjugate. The normalized autocovariance function is called the autocorrelation function. However, Wiener generally used the term autocorrelation for $\phi(k)$, whether it was normalized or not. Nevertheless, it is confusing to keep using the term "autocorrelation" with two different meanings. It is better to use the term "autocovariance" wherever it is appropriate.

A white noise process is stationary. In the case of continuous time, its autocorrelation is $\delta(t)$ (the Dirac delta), whereas in the case of discrete time, its autocorrelation is δ_k (the Kronecker delta). As we have seen, the Fourier transform $E(\omega)$ of white noise in time is white in frequency; that is, the autocorrelation in frequency is the Dirac delta function

$$E E^*(\omega) E(\omega + \mu) = 2\pi\delta(\mu).$$

The problems confronting empirical workers in spectral analysis in the first part of the twentieth century were centered around the Schuster periodogram. Schuster introduced this concept at the turn of the century, and until Yule's work in 1927, it was the only method available to carry out empirical spectral analysis. Suppose that we observe a stationary random process for a very long time, so that we obtain a time series $x(n)$ for $n = 1, 2, \cdots, N$, where N is very large. Schuster then computed the periodogram

$$P(\omega) = \frac{1}{N} |X(\omega)|^2$$

where $X(\omega)$ is the discrete Fourier transform

$$X(\omega) = \sum_{n=1}^{N} x(n) e^{-j\omega n}.$$

(Today we can compute $X(\omega)$ very rapidly by means of the Cooley–Tukey fast Fourier transform, but then it was a formidable task.) In the case when the stationary process is made up of sinusoidal waves with superimposed white noise, the periodogram is effective in picking out the discrete frequencies of the sinusoids. But a purely nondeterministic stationary process is generated by the convolution formula (input–output relation)

$$x(n) = \sum_{k=0}^{\infty} b(k)\, \epsilon(n - k).$$

(Here we interpret $b(k)$ as the impulse response function of a filter, the white noise process $\epsilon(n)$ as the input to the filter, and the stationary process $x(n)$ as the output). For such a process, the Schuster periodogram $P(\omega)$ is extremely rough, and often cannot readily be interpreted. Empirical spectral analysis was at an impasse. Most of the time series observed in nature could not be analyzed by the methods available in 1930.

Now comes Wiener in 1930 with generalized harmonic analysis. In brief, Wiener in 1930 knew how to take the Fourier transform of a stationary random process, a milestone in the use of Fourier methods. Wiener's generalized harmonic analysis makes use of a generalized random function, namely, the Einstein–Wiener (white noise) process. In order to put Wiener's work into context, we will now give a small digression on the most widely known generalized function: the Dirac delta function.

The *impulse* (*Dirac delta*) *function* had been known for many years prior to its use by Dirac [26] in 1928. It was known by Heaviside [27]. However, it took the stature of a great physicist, Paul Dirac, to decree in 1928 the use of the impulse function in physics. In those early days, people used to talk about $\delta(t)$ as a function of t in the ordinary sense whose integral with $f(t)$ produces $f(0)$; that is,

$$\int_{-\infty}^{\infty} \delta(t) f(t)\, dt = f(0).$$

This idea used to cause great distress to mathematicians, some of whom even declared that Dirac was wrong despite the fact that he kept getting consistent and useful results. The physicists rejected these extreme criticisms and followed their intuition. We can now see why the physicists succeeded despite the reservations of the mathematicians. It is true that the physicists spoke of $\delta(t)$ as an ordinary function, which it cannot be in any precise sense, and that they treated it as an ordinary function by integrating it and even differentiating it. But the physicists were justified because they only used $\delta(t)$ inside integrals with sufficiently-differentiable functions $f(t)$. For example, the derivative of $\delta(t)$ always appeared inside an integral, and the integral was integrated by parts as follows:

$$\int_{-\infty}^{\infty} \delta'(t) f(t)\, dt = -\int_{-\infty}^{\infty} \delta(t) f'(t)\, dt = -f'(0).$$

The physicists never used the delta function except to map functions to real numbers. In this sense, they employed the machinery but not the words of distribution theory, which was devised expressly in order to give delta functions a sound basis. It was the French mathematician L. Schwartz who after World War II created a systematic theory of generalized functions and explained it in his well-known monograph *Théorie des Distributions* in 1950 and 1951. From then on the theory of generalized functions was developed intensively by many mathematicians. This precipitate development of distribution theory received its main stimulus from the requirements of mathematics and theoretical physics, in particular the theory of differential equations and quantum physics. Generalized functions possess a number of remarkable properties that extend the capabilities of classical mathematical analysis. For example, any generalized function turns out to be infinitely differentiable (in the generalized meaning), convergent series of generalized functions may be differentiated termwise an infinite number of times, the Fourier transform of a generalized function always exists, and so on. For this reason, the uses of generalized function techniques substantially expand the range of problems that can be tackled and leads to appreciable simplifications that make otherwise difficult operations automatic.

As science advances, its theoretical statements seem to require an ever higher level of mathematics. When he gave his theoretical prediction of the existence of antiparticles in 1931 (*Proc. Roy. Soc. London, Ser.* A, vol. 133, pp. 60–72) Dirac wrote, "It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalization of the axioms at the base of mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation." Subsequent developments in theoretical physics have corroborated this view. In this essay, we have seen that since the time of Newton, the search for and the study of mathematical models of physical phenomena have made it necessary to resort to a wide range of mathematical tools and have thus stimulated the development of various areas of mathematics. Now let us return to Norbert Wiener in 1930.

Physicists are concerned with unlocking the mysteries of nature, and the impulse (Dirac delta) function eases their task. The impulse function is the simplest of the generalized functions. One can imagine the plight of Wiener in the mathematical community when he introduced generalized random functions into the mathematical literature as early as 1923, and especially in his 1930 paper.

Let us now give the gist of Wiener's *generalized harmonic analysis*. As we know, convolution in the time domain corresponds to multiplication in the frequency domain. Thus in terms of Fourier transforms, the above input–output convolu-

tion integral becomes

$$X(\omega) = B(\omega)\, E(\omega).$$

In this equation, $E(\omega)$ is the Fourier transform of white noise, so that $E(\omega)$ is a generalized function that is white (i.e., very rough) in frequency. The filter's transfer function $B(\omega)$ is a smooth well-behaved (ordinary) function. The product $X(\omega)$ is also very rough.

Let us now take the inverse Fourier transform of $X(\omega)$. It is

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n} X(\omega)\, d\omega$$

which is

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n} B(\omega)\, E(\omega)\, d\omega.$$

This formula represents Wiener's generalized harmonic analysis of $x(n)$; that is, it is the *spectral representation* of the stationary random process $x(n)$. It involves the smooth filter transfer function $B(\omega)$ and the very rough (white in frequency) process $E(\omega)$. We thus see that the spectral representation requires Wiener's generalized random function $E(\omega)$, which came out of his studies of Brownian movement.

Wiener's generalized harmonic analysis (i.e., spectral representation) explains why the periodogram of Schuster did not work for convolutional processes. Because the periodogram (as the number of observations becomes large) is

$$P(\omega) = \frac{1}{N}|X(\omega)|^2$$

it follows that the periodogram has the intrinsic roughness of the $X(\omega)$ process. (It was not until the work of J. Tukey [34] in 1949 that a means was found to overcome this problem; Tukey's breakthrough was of epoch proportions.)

Wiener in his 1930 paper gave the following method, which was standard until the work of Tukey in 1949. Wiener's method was intended for very long time series. It consisted of computing the autocovariance function as the time average

$$\phi(k) = \frac{1}{N}\sum_{n} x^*(n)\, x(n+k)$$

for $-p \leqslant k \leqslant p$, where p is less than the data length N, and then computing the power spectrum $\Phi(\omega)$ as the Fourier transform

$$\Phi(\omega) = \sum_{k=-p}^{p} \phi(k)\, e^{-j\omega k}.$$

This Fourier transform relationship between autocovariance and power spectrum, as we have observed, is now called the *Wiener–Khinchin* theorem.

Whereas von Neumann's work in quantum physics in 1929 received instant acclaim and well-deserved recognition by physicists and mathematicians, Wiener's work in 1930 lay dormant. However, now with the benefit of hindsight, it is worthwhile for us to reconcile these two approaches to spectral estimation. This we will do in the next section.

XI. Reconciliation of the Two Spectral Theories

We have come a long way in the history of spectral estimation to this point. From the work of the ancients in deriving a calendar, to the work of the great mathematicians who formulated the wave equation in the eighteenth century, it took thousands of years. Then the work of Bernoulli, Euler, and Fourier came, and the result was a spectral theory in terms of sinusoidal functions, in place at the beginning of the nineteenth century. The theory was extended to the case of arbitrary orthogonal functions by Sturm and Liouville, and this led to the greatest empirical success of spectral analysis yet obtained: the physical results of spectral estimation that unlocked the secret of the atom. Credit for this result belongs to Heisenberg and Schrödinger in 1925 and 1926. Then in 1929, the work of von Neumann put the spectral theory of the atom on a firm mathematical foundation in his spectral representation theorem. The spectral work of von Neumann represents the cumulation of this line of research in quantum physics. Meanwhile, Rayleigh and Schuster at the beginning of the twentieth century were applying the original sinusoidal methods of Fourier to the analysis of data in the realm of classical physics. However, the periodogram approach of Schuster did not work well for purely nondeterministic stationary random processes, and this led Yule in 1927 to develop a spectral theory for a subclass known as autoregressive processes. Meanwhile, Wiener had developed the mathematical theory of Brownian movement in 1923, and in 1930 he introduced generalized harmonic analysis, that is, the spectral representation of a stationary random process. Thus in 1930, we have two spectral theories, one represented by the spectral representation theorem of von Neumann and the other by the spectral representation theorem of Wiener. It is the purpose of this section, with the benefit of hindsight, of course, to indicate the relationship between von Neumann and Wiener spectral theories.

The common ground is the Hilbert space. As we have seen, the von Neumann result is the spectral representation of a Hermitian operator H in Hilbert space. The Schrödinger equation is written in terms of a Hermitian operator, and this equation governs the spectrum of atoms and molecules. Now let us, however, leave this Hilbert space and look at another one. The other Hilbert space is one defined by the probability measure that governs the stationary random process in question. As we know, a Hilbert space is specified by an inner (or dot) product. The elements of the Hilbert space are random variables, and the inner product is defined as the expected value given by

$$\langle x, y\rangle = Ex^*y.$$

(The superscript asterisk indicates the complex conjugate.) In this Hilbert space, a stationary process is defined as follows. We use discrete (integer) time n, although a similar development can be made in the case of continuous time. A sequence of random variables $x(n)$ in Hilbert space is called a *stationary random process* if its autocorrelation

$$\phi(k) = \langle x(n), x(n+k)\rangle$$

depends only upon the time-shift k and not on absolute time n. This definition implies that the elements $x(n)$ of the process are generated recursively by a unitary operator; that is,

$$Ux(n) = x(n+1)$$

so that

$$x(n+k) = U^k x(n).$$

Because a unitary operator represents a rotation, we see that a stationary random process traces out a spiral in Hilbert space, the so-called *Wiener spiral*. We now come to the connection

that we are seeking, namely, the fact that the Cayley-Möbius transformation [28] of a Hermitian operator is a unitary operator. Thus there is a one-to-one correspondence between Hermitian operators and unitary operators in Hilbert space. The von Neumann spectral representation is for a Hermitian operator. If we take its Cayley-Möbius transformation, we obtain the corresponding spectral representation for the unitary operator U. This spectral representation has the form

$$U = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega}\mathfrak{U}(\omega)\,d\omega$$

where $\mathfrak{U}(\omega)$ represents a family of projection operators as a function of circular frequency ω. Thus the process has the representation

$$x(n) = U^n x(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n}\mathfrak{U}(\omega)\,x(0)\,d\omega.$$

We now make the identification

$$\mathfrak{U}(\omega)\,x(0) = X(\omega)$$

and we obtain

$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega n}X(\omega)\,d\omega.$$

This equation is Wiener's generalized harmonic analysis of the process. Thus we have the connection we sought; the two spectral representations are related by the Cayley-Möbius transformation.

XII. Wiener-Levinson Prediction Theory

Early in 1940, Wiener became involved in defense work at MIT and, in particular, he became interested in the design of fire-control apparatus for anti-aircraft guns. The problem was to build into the control system of the gun some mechanical device to aim the gun automatically. The problem, in effect, was made up of two parts: a mathematical part, which consisted of predicting the future position of an airplane from its observed past positions, and an engineering part, which consisted of realizing the mathematical solution in the form of an actual physical device. Wiener recognized that it was not possible to develop a perfect universal predictor, and so he formulated the mathematical problem on a statistical basis. He defined the optimum predictor as the one that minimizes the mean-square prediction error. The minimization led to the Wiener-Hopf integral equation, which represented the completion of the mathematical part of the problem. As to the engineering part, Wiener immediately recognized that it was possible to devise a hardware apparatus that represents the solution to the Wiener-Hopf equation. As Wiener [29] states in his autobiography (p. 245): "It was not hard to devise apparatus to realize in the metal what we had figured out on paper. All that we had to do was make a quite simple assembly of electric inductances, voltage resistances, and capacitors, acting on a small electric motor of the sort which you can buy from any instrument company." Wiener's mathematical results [24] were published in 1942 as a classified report to Section D2 of the National Defense Research Committee. This report is Wiener's famous *Time Series* book, which we mentioned previously. Its full title is *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, and it was republished as an unrestricted document in 1949 by MIT Press, Cambridge.

Although Wiener's "General Harmonic Analysis" did not have immediate influence, his *Time Series* book, which was written in a more understandable style, did among those who had access to the book in 1942 and the general public in 1949. As we will now see, a great deal of credit for the dissemination of Wiener's ideas belongs to his former student and his colleague, Professor N. Levinson.

Levinson's initial contact with Wiener was in Wiener's course in 1933-34 on Fourier Series and Integrals, which is described in Levinson's own words as follows:

> "I became acquainted with Wiener in September 1933 while still an undergraduate student of electrical engineering, when I enrolled in his graduate course. It was at that time really a seminar course. At that level he was a most stimulating teacher. He would actually carry out his research at the blackboard. As soon as I displayed a slight comprehension of what he was doing, he handed me the manuscript of Paley-Wiener for revision. I found a gap in a proof and proved a lemma to set it right. Wiener thereupon set down at his typewriter, typed my lemma, affixed my name and sent it off to a journal. A prominent professor does not often act as a secretary for a young student."

N. Levinson, a dynamic and brilliant mathematician and a warm and kind person, made important and permanent contributions to engineering and applied science. The Levinson theorem in quantum mechanics illustrates his ability to grasp the relationship between physical concepts and mathematical structure. Few have this insight, and nowhere is it better demonstrated than in the two expository papers written in 1942 by Levinson soon after the restricted publication of Wiener's *Time Series* book. These two papers were published in 1947 in the *Journal of Mathematical Physics*, and thus they represented the first public disclosure of Wiener's time series results. Later these two papers also appeared as Appendices C and B in the unrestricted publication of Wiener's book in 1949 by MIT Press [24].

An appreciation of Levinson's contribution can be gained in historical perspective. The 1942 edition of Wiener's book was bound in a yellow paper cover, and because of its difficult mathematics, it came to be known among engineers as the "yellow peril" (a term familiar to mathematicians as applying to a famous series of advanced texts). However clear in a conceptual way the building of an actual device was to Wiener, there were few engineers at that time who were able to grasp Wiener's mathematical solution, much less to realize it in the form of a physical device. At this point, Levinson stepped in and wrote "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," [30], one of his two classic applied papers on explaining Wiener's work. Levinson describes his paper as an expository account of Wiener's theory. Levinson's earlier training was in electrical engineering, so he understood hardware design methods. In the paper, Levinson shows in an elementary way why the Wiener-Hopf equation cannot be solved by use of the Fourier transform theorem. Then, in a natural way, he introduces the spectral factorization and obtains the explicit solution for the prediction operator and, more generally, for the filter operator. This masterpiece of exposition opened up these methods to the engineering profession.

Levinson's other classic applied paper is entitled, "The Wiener RMS (root mean square) error criterion in filter design and prediction." [31] As before, let us try to put this paper in historical perspective. In 1942, the Army Air Force Weather Division negotiated a contract with MIT to perform statistical

analyses of meteorological and climatological data, particularly in relationship to weather forecasting, and to conduct research into the application of statistical techniques to long-range forecasting [32]. Professor G. P. Wadsworth of the MIT Mathematics Department was in charge of this meteorological project. The basic idea was to collect and sort large amounts of numerical meteorological data and to forecast by analogy, much like the forecasts made on television today, in which the weatherman looks at the data appearing on the satellite picture of the earth. Wadsworth's method had merit, but the data required was just not available in the 1940's. Wiener's *Time Series* book was completed at about the same time as this MIT Meteorological Project was starting up. Since the weather data available occurred at discrete intervals of time, the continuous-time methods of Wiener were not directly applicable. As a result, Wadsworth asked Levinson to write up a discrete form of Wiener's theory. The result was Levinson's "Wiener RMS" paper with the Levinson recursion. However, use was never made of Wiener–Levinson prediction theory by the MIT Meteorological Project, and Levinson's paper sat dormant.

In order to understand why Levinson's methods were not used in the 1940's, one must look at the computing facilities available at the time. The actual realization of these methods would have to be carried out by people using hand calculators. A hand calculator could add, subtract, multiply, and divide, but had no memory except an accumulator. Thus the result of each separate calculation had to be transferred to paper by hand, a drawn-out, time-consuming process. In contrast to the hardware devices working in real time, as envisaged by Wiener, the hand calculator was a poor substitute. As Wiener [24, p. 102] states: "Much less important, though of real interest, is the problem of the numerical filter for statistical work, as contrasted with the filter as a physically active piece of engineering apparatus."

After Levinson wrote his two expository papers, which were completed in 1942, neither he nor Wiener took up research in the computational (software) aspect of Wiener's theory. Wiener was more interested in its realization by machines (hardware), and his research interest was already shifting to biological and medical problems. In fact, it was the union of these two research interests that led to his discovery and formulation of the science of cybernetics, which he describes as the problem of control and communication in machines and animals. Meanwhile Levinson had decided as early as 1940 to shift his field from the Fourier methods of Wiener to the field of nonlinear differential equations. He talked about this decision with his friends in 1940. Levinson worked hard over a period of two or three years (which included the period during which he wrote the two expository papers) before he felt that he had enough mastery in his new field. Such mastery he did achieve, and his outstanding contributions to differential equations were recognized by his receiving the prestigious Bocher Prize in Mathematics in 1954.

Despite their other research interests, both Wiener and Levinson were always ready to give their support and time to the MIT Meteorological Project directed by G. P. Wadsworth. Wiener was especially interested in seeing physical examples of autocorrelation functions. This interest led to the computation of several autocorrelation functions of ocean wave data by Wadsworth and by his friend and associate H. R. Seiwell [33], who was with the Woods Hole Oceanographic Institution. The interest in these computations led to the "Symposium on Autocorrelation Analysis Applied to Physical Problems" held at Woods Hole, MA in June 1949, sponsored by the Office of Naval Research. The high point of this meeting was the paper by Tukey [34]. Before Tukey's work, the power spectra computed from empirical autocorrelation functions were too erratic to be of any use in formulating physical hypotheses. Not only did Tukey show correctly how to compute power spectra from empirical data, but he also laid the statistical framework for the analysis of short-time series, as opposed to the very long ones envisaged by Wiener and Levinson.

Wadsworth was also director of the MIT section of the U.S. Naval Operations Evaluation Group, a project started in World War II which initiated the use of operations research in the United States. By 1950, Wadsworth was applying operations research methods to industry and had established himself as one of the highest paid consultants in the United States. There were so many industrial people waiting to see him in his outer office at MIT that one had to make an appointment with his secretary many weeks in advance to see him in his inner office for just 5 or 10 minutes at the most. The writer began as one of Wadsworth's research assistants in the MIT Mathematics Department in September 1950, and he was assigned to work in seismology by Professor Wadsworth. Mobil Oil made available eight seismic records, and the writer immediately got a very lonely feeling, especially at MIT at night digitizing the Mobil seismic records with a ruler and pencil. Except for Wadsworth, in 1950 nobody at MIT or in the oil industry thought that the analysis of digital seismic data would ever be feasible.

Fortunately Tukey took an interest in the seismic project and conveyed his research ideas by mail. The first empirical results were the computation of the Tukey spectra for various sections of the Mobil records in the spring of 1951. From these spectral results, a seismic analysis based on prediction error was formulated in the summer of 1951. This analysis made use of Wiener prediction theory in digital form. Prior to this work, Wiener's procedures had only been realized in analog form. In hand plotting the first numerical results of what today is called linear predictive coding (LPC), the writer was so amazed that he could not believe his eyes, and he was sure that he would never see such good results again. But the second trace, and the third trace, and so on, were computed and confirmed what he saw. The digital processing method called deconvolution worked! As soon as possible, he made an appointment to see Wadsworth, which the secretary set three weeks from then, in September 1951. The result was that the digitally-processed seismic traces [35] were sent out to the oil industry, and the oil companies gave money to support a project. The MIT Geophysical Analysis Group was thus born, and the MIT Whirlwind digital computer was used to analyze seismic records throughout the lifetime of the GAG (1952–1957). During this period, Tukey freely gave his research advice [36], [37]. For example, Tukey's methods for estimating coherency (today called by various names, such as "semblance" by the oil industry) are vital in the estimation of seismic velocity as well as in other multichannel methods. Tukey's vision of a fast Fourier transform was always influential. In fact, S. M. Simpson [38], who later directed the Geophysical Analysis Group, eventually devised an efficient 24-point Fourier transform, which was a precursor to the Cooley–Tukey fast Fourier transform in 1965. The FFT made all of Simpson's efficient autocorrelation and spectrum programs instantly obsolete, on which he had worked the

equivalent of half a lifetime. Wiener was very generous of his time [39]. Wiener's work [40], [41] on multichannel methods was helpful later in extending the Levinson recursion, which Levinson had devised for single channel time series, to the multichannel case [42]. The excellent seismic data and corresponding well logs supplied by the oil industry to the Geophysical Analysis Group made possible the development of the statistical minimum-delay model of the earth's stratigraphic layers, together with the theoretical justification of seismic deconvolution [19], [43].

In the late 1950's a digital revolution occurred because of the introduction of transistors in the building of digital computers, which made possible reliable computers at a much lower cost than previously. As a result, the seismic industry completely converted to digital technology in the early 1960's, a long ten years after the first digital results were obtained. Since then, nearly every seismic record taken in the exploration of oil and natural gas has been digitally deconvolved and otherwise digitally processed by these methods. The final result of the digital processing of seismic data was the discovery of great oil fields which could not be found by analog methods. These oil fields include most of the offshore discoveries, as in the North Sea, the Gulf of Mexico, the Persian Gulf, as well as great onshore discoveries in Alaska, Asia, Africa, Latin America, and the Middle East, made in the last twenty years. Today an oil company will deconvolve and process as many as one million seismic traces per day; it took a whole summer in 1951 to do 32 traces.

Whereas the digital revolution came first to the geophysical industry largely because of the tremendous accuracy and flexibility afforded by large digital computers, today we are in the midst of a universal digital revolution of epic proportions. One now realizes that the work of Wiener and Levinson is being appreciated and used by an ever-increasing number of people. Digital signal processing is a growing and dynamic field which involves the exploration of new technology and the application of the techniques to new fields. The technology has advanced from discrete semiconductor components to very large-scale integration (VLSI) with densities above 100 000 components per silicon chip. The availability of fast, low-cost microprocessors and custom high-density integrated circuits means that increasingly difficult and complex mathematical methods can be reduced to hardware devices as originally envisaged by Wiener, except the devices are digital instead of analog. For example, a custom VLSI implementation of linear predictive coding is now possible, requiring a small number of custom chips. Whereas originally digital methods were used at great expense only because the application demanded high flexibility and accuracy, we have now reached the point that anticipated long-term cost advantages have become a significant factor for the use of digital rather than analog methods.

XIII. Tukey Empirical Spectral Analysis

As we have mentioned, a turning point in the empirical analysis of time series data began in 1949 at the Woods Hole Symposium on Applications of Autocorrelation Analysis. There Tukey presented the first of three papers [34], [36], [37], which he had written in the early years on spectrum analysis. These papers introduced the classic Tukey method of numerical spectral estimation, a method that has been used by most workers since that time. In addition, Tukey described an approximate distribution for the estimate. This distribution was required for the proper design of experiments for the collection of time series data. In a very interesting paper [44], Tukey describes the situation which led to his spectral work, including a discussion of Hamming's suggestion about the smoothing of the discrete Fourier transform of an empirical autocorrelation, which led to the joint work of Hamming and Tukey.

During the last four decades, Tukey [45]–[50] introduced a multitude of terms and techniques that are standard to the practice of the data analysis of time series. Such commonplace terms and concepts as "prewhitening," "aliasing," "smoothing and decimation," "tapering," "bispectrum," "complex demodulation," and "cepstrum" are due to Tukey. Very few papers in the literature of applied time series analysis do not give some acknowledgment of Tukey's ideas and methods, and most papers credit his ideas in some vital way. Moreover, Tukey [51]–[55] has made substantial contributions in the placing of the data analysis of time series into perspective with current research in the physical sciences, in statistics, and in computing and numerical analysis.

We have already mentioned the key influence Tukey had on the MIT group, which included Wadsworth, Simpson, the writer, and others. W. J. Pierson and L. J. Tick [56] at New York University used Tukey's methods in the analysis of oceanographic time series records. The outstanding thesis of Goodman [57], which extended the results of Tukey to the bivariate case, was written under Tukey's supervision. The group at La Jolla, CA, which included Munk, Rudnick, and Snodgrass, applied Tukey's spectral methods to estimate wave motion due to storms many thousands of miles away, a testimony to the power of his methods. Munk and McDonald wrote a remarkable book, *The Rotation of the Earth* [58], which used these spectral methods in several novel ways.

In this period, packages of computer programs for time series analysis were appearing. The collection of programs by Healy [59] of Bell Laboratories were circulated from 1960 on. The BOMM collection of programs [60] was developed at La Jolla. Some of the programs used by Parzen were included in his book [61]. The programs written at MIT are described in [38] and [62].

In econometrics, Granger's book [63] in 1964 described many of the techniques suggested by Tukey for the analysis of univariate and bivariate time series. The most successful application of spectrum techniques to economic series is its use for the description of the multitude of procedures of seasonal adjustment. In astronomy, Neyman and Scott [64] in 1958 carried out the analysis of two-dimensional data consisting of the positions of the images of galaxies on photographic plates.

Norbert Wiener remained active until his death in 1964. His later work included both empirical results, such as modeling and analyzing brain waves [66], [67], and theoretical results, such as his work with Masani on multivariate prediction theory [40], [41]. Wiener's death marks the end of an era in time series analysis and spectral theory.

XIV. The Cooley-Tukey Fast Fourier Transform

The present epoch of time series analysis began in 1965 with the publication of the fast Fourier transform by Cooley and Tukey [68]. The effect that this paper has had on scientific and engineering practice cannot be overstated. The paper described an algorithm for the discrete Fourier transform of $T = T_1 \cdots T_p$ values by means of $T(T_1 + \cdots + T_p)$ multi-

plications instead of the naive number T^2. Although such algorithms existed previously [69], they seem not to have been put to much use. Sande developed a distinct, symmetrically related algorithm simultaneously and independently.

The existence of such an algorithm meant, for example, that the following things could be computed an order of magnitude more rapidly: spectrum estimates, correlograms, filtered versions of series, complex demodulates, and Laplace transforms (see, for example, [70]). General discussions of the uses and importance of fast Fourier transform algorithms may be found in [71] and [72]. The Fourier transform of an observed stretch of series can now be taken as a basic statistic and classical statistical analyses—such as multiple regression, analysis of variance, principal components, canonical analysis, errors in variables, and discrimination—can be meaningfully applied to its values, [73] and references cited therein. Higher-order spectra may be computed practically [74]. Inexpensive portable computers for carrying out spectral analysis have appeared on the market and may be found in many small laboratories.

The years since 1965 have been characterized by the knowledge that there are fast Fourier transform algorithms. They have also been characterized by the rapid spread of type of data analyzed. Previously, the data analyzed consisted almost totally of discrete or continuous real-valued time series. Now the joint analysis of many series, such as the 625 recorded by the Large Aperture Seismic Array in Montana [75], has become common. Spatial series are analyzed [75]. The statistical analysis of point processes has grown into an entirely separate field [76]. The SASE IV computer program developed by Peter Lewis [77] has furthered such analysis considerably. We note that transforms other than the Fourier are finding interest as well [78], [79].

XV. Burg Maximum Entropy Spectral Analysis

The discrete Fourier transform of the autocovariance is called the power spectrum. Thus the power spectrum of a stationary time series with autocovariance $\phi(n)$ is

$$\Phi(\omega) = \sum_{n=-\infty}^{\infty} \phi(n)\, e^{-j\omega n} \quad (\text{for } -\pi \leqslant \omega \leqslant \pi). \qquad (1)$$

In this discussion, we do not require the autocovariance to be normalized, so $\phi(0)$ does not have to be one. If we know the entire autocovariance function, that is, if we know $\phi(n)$ for all values of n, then, of course, we can obtain the power spectrum by means of (1). However, in many applications, we know or can reliably measure the autocovariance only for a certain finite number of values of n, say for $n = 0, 1, 2, \cdots, p$. Because the autocovariance is symmetrical $\phi(n) = \phi(-n)$, we thus know the values of $\phi(n)$ for $n = 0, \pm 1, \pm 2, \cdots, \pm p$, and we do not know $\phi(n)$ for $|n| > p$. The question is how should we estimate the power spectrum (1) from only this partial knowledge.

In order to answer this question, we should first consider the phenomenon under study. As there are many different types of phenomena, there is a corresponding diversity in spectrum analysis. Brillinger and Tukey state that no distinction is more vital than that among a) noise-like processes, b) signal-like processes, and c) signal-plus-noise processes. While b) is ordinarily unrealistic, there are enough cases of c) with only a slight amount of noise so that b) is a helpful idealization.

A noise-like process produces time series quite different in character than those produced by a signal-like process. The regularity of a noise-like process does not lie in the shapes of its individual realizations, but in its underlying statistical structure. As a result, different realizations usually do not appear to resemble each other. A segment of one realization generally will neither look the same nor have the same empirical autocovariances as the corresponding segment of another realization. Anyone using the autocovariance from a given realization as a "known" autocovariance, and then fitting this "known" autocovariance exactly, is likely to be in worse error than if he does not. As Brillinger and Tukey point out, we often have only one distinct realization, and we need to make as good inferences as we can about the underlying population. We have to think statistically, and treat our numerical results with due consideration as to the tentative nature of the structural and stochastic assumptions built into the model. It is for data near the noise-process end of the continuum that many of the Fourier methods of spectrum estimation are intended and are most effective.

At the other extreme are the signal-like processes. Some of the best examples of such processes are found in exploration seismology. A concentrated energy source produces a seismic record. Because the random background noise is usually weak, the record will appear essentially the same as another one taken at the same place at a different time. As Brillinger and Tukey point out, the study of signal-like processes may well be done by quite different methods.

Also there is another aspect of spectrum analysis, one to which the limitations of the data have forced many applied research workers. Here the models are not narrowly restricted by reliable subject-matter knowledge. The time-series records are not long, and the appearance of the data is not distinctive. Corresponding records do not look alike. With almost nothing to work with, the research workers can do nothing much more than fitting a few constants. Thus they fit low-order AR, MA, and ARMA models usings the Box–Jenkins [93] methodology. As Brillinger and Tukey observe, this approach in a large number of applications seems to work much better than might be anticipated.

The basic issue in spectrum estimation is the proper choice of model and then the resulting choice of the method of spectrum estimation [102]. Application of a particular spectral estimator to an inappropriate model can result in serious specification errors.

Let us now return to the question in point, namely, how we should estimate the power spectrum from a limited section of the autocovariance. Most research workers who calculate and use only a limited number of autocovariances do not assume that all the later ones vanish. They use the earlier values to calculate a quadratic function of the data whose average value (across the ensemble), like all other quadratic functions of the data, is the integral of a knowable kernel with the spectrum. They then interpret their result accordingly. One must not charge these workers with making any assumption about the uncalculated autocovariances. What they usually do is to recognize that the empirical autocovariances that they do calculate will vary from realization to realization, and so they do not and should not take the empirical values as absolute truth.

One of the purposes of this historical essay is to try to give the flavor of important developments in spectrum estimation. Unfortunately, the writer did not attend the 37th Meeting of the Society of Exploration Geophysicists in 1967, although he attended the meetings in the years both before and after that

one. It was at that meeting in Oklahoma City that John Burg presented a paper that was to shake the foundations of spectrum estimation. This fundamental work [80] is entitled *Maximum Entropy Spectral Analysis* and its abstract reads:

> "The usual digital method of obtaining a power spectrum estimate from a measured autocovariance function makes the assumption that the correlation function is zero at all lags for which no estimate is available and uses some treatment of the estimated lags to reduce the effect of truncation of the autocovariance function. The method discussed in this paper instead retains all of the estimated lags without modification and uses a nonzero estimate for the lags not directly estimated. The particular estimation principle used is that the spectral estimate must be the most random or have the maximum entropy of any power spectrum which is consistent with the measured data. This new analysis technique gives a much higher resolution spectral estimate than is obtained by conventional techniques with a very little increase in computing time. Comparisons will illustrate the relative importance."

The estimation of the power spectrum of stationary time series from partial knowledge of its autocovariance function is a classical problem to which much attention has been given over the years. Almost all of this work is based on the use of window functions, whose properties can be analyzed by Fourier methods. Burg [80], [81], [82] in his pioneering work introduced a new philosophy in spectral analysis based on general variational principles, and, in particular, the *maximum entropy method* (MEM) which we will now discuss.

The conventional approach to estimating the power spectrum from $\phi(n)$, $|n| \leqslant p$, is to assume that $\phi(n) = 0$ for $n > p$ and to take the Fourier transform of $w(n)\,\phi(n)$, $|n| \leqslant p$, where $w(n)$ is a weighting function. We now want to describe the maximum entropy method of Burg.

Given a limited set of autocovariance coefficients together with the fact that a power spectrum $\Phi(\omega)$ must be non-negative, we know that there are generally an infinite number of power spectra in agreement with this information. Thus additional information is required, and a reasonable goal is to find a single function $\Phi(\omega)$, which is representative of the class of all possible spectra. In order to resolve this problem, some choice has to be made, and Burg made use of the concept of entropy. Maximum entropy spectral analysis is based on choosing that spectrum which corresponds to the most random or the most unpredictable time-series whose autocovariance coincides with the given set of values. This concept of maximum entropy is the same as that used in both statistical mechanics and information theory, and as we will see represents the most noncommittal assumption possible with regard to the unknown values of the autocovariance function.

Equation (1) gives the power spectrum $\Phi(\omega)$ as the discrete Fourier transform of the autocovariance function $\phi(n)$. From Fourier theory, we know that the autocovariance function can be obtained as the inverse Fourier transform of the power spectrum; that is

$$\phi(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\omega)\, e^{j\omega n}\, d\omega \qquad \text{(for all integers } n\text{)}. \quad (2)$$

The fundamental assumption involved in maximum entropy spectral analysis is that the stationary process under consideration is the most random or the least predictable time series that is consistent with the given measurements. Specifically, the given measurements are the known autocovariance coefficients, namely

$$\phi(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\omega)\, e^{j\omega n}\, d\omega \qquad (\text{for } |n| \leqslant p). \quad (3)$$

In terms of information theory, we require that the entropy per sample of time series is a maximum. From the work of Shannon in 1948, it follows that the entropy is proportional to the integral of the logarithm of the power spectrum, that is, the entropy is

$$\int_{-\pi}^{\pi} \log \Phi(\omega)\, d\omega. \quad (4)$$

Therefore the required maximum entropy power spectrum is that function $\Phi(\omega)$ which maximizes (4) under the constraint equations (3).

One way to solve the problem of finding the maximum entropy power spectrum subject to fixed values of $\phi(n)$ for $|n| \leqslant p$ is by use of Lagrange multipliers. However, we may instead use the following approach. From (1) we see that the partial derivative of $\Phi(\omega)$ with respect to $\phi(n)$ is

$$\frac{\partial \Phi(\omega)}{\partial \phi(n)} = e^{-j\omega n}.$$

It follows that

$$\frac{\partial \log \Phi(\omega)}{\partial \phi(n)} = \frac{e^{-j\omega n}}{\Phi(\omega)} = [\Phi(\omega)]^{-1} e^{-j\omega n}. \quad (5)$$

Now let us maximize (4) with respect to the unknown values $\phi(n)$ where $|n| > p$. Thus we set the partial derivatives of (4) with respect to $\phi(n)$ for $|n| > p$ equal to zero; that is

$$\frac{\partial}{\partial \phi(n)} \int_{-\pi}^{\pi} \log \Phi(\omega)\, d\omega = \int_{-\pi}^{\pi} \frac{\partial \log \Phi(\omega)}{\partial \phi(n)}\, d\omega = 0, \qquad \text{for } |n| > p. \quad (6)$$

Making use of (5), we see that (6) reduces to

$$\int_{-\pi}^{\pi} [\Phi(\omega)]^{-1} e^{-j\omega n}\, d\omega = 0, \quad \text{for } |n| > p. \quad (7)$$

This equation specifies the form of the inverse power spectrum $[\Phi(\omega)]^{-1}$ of a maximum entropy process. Let us explain.

Given any stationary process with positive power spectrum $\Phi(\omega)$, then its inverse power spectrum

$$[\Phi(\omega)]^{-1} = \frac{1}{\Phi(\omega)} \quad (8)$$

is also positive. We assume that the inverse power spectrum is integrable and bounded, so that it is a well-behaved power spectrum in its own right. Thus the inverse power spectrum (8) can be associated with an autocovariance function, which we designate by $\psi(n)$, such that the counterparts of (1) and (2) hold, namely

$$[\Phi(\omega)]^{-1} = \sum_{n=-\infty}^{\infty} \psi(n)\, e^{-j\omega n} \qquad (\text{for } -\pi \leqslant \omega \leqslant \pi) \quad (9)$$

and

$$\psi(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} [\Phi(\omega)]^{-1} e^{j\omega n}\, d\omega \qquad \text{(for all integers } n\text{)}. \tag{10}$$

Because we do not normalize, the zero-lag value $\psi(0)$ does not have to be one.

Let us now return to the maximum entropy process. As we have seen, its inverse power spectrum satisfies (7) for $|n| > p$. In (7) replace n by $-n$, and also multiply each side of the equation by $1/2\pi$. Thus we obtain

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} [\Phi(\omega)]^{-1} e^{j\omega n}\, d\omega = 0, \quad \text{for } |n| > N. \tag{11}$$

Comparing (11) with (10), we see that for a maximum entropy process we have

$$\psi(n) = 0, \quad \text{for } |n| > p. \tag{12}$$

Hence, using (12) in (9), we see that the inverse power spectrum of a maximum entropy process is

$$[\Phi(\omega)]^{-1} = \sum_{n=-p}^{p} \psi(n)\, e^{-j\omega n}. \tag{13}$$

The right-hand side of (13) is a finite trigonometric series. Thus we have shown that the maximum entropy process is one whose inverse power spectrum is a finite trigonometric series, or equivalently one whose power spectrum is the reciprocal of a finite trigonometric series, that is

$$\Phi(\omega) = \frac{1}{\sum_{n=-p}^{p} \psi(n)\, e^{-j\omega n}}. \tag{14}$$

If we let $z = e^{j\omega}$, then the finite trigonometric series (13) becomes

$$\frac{1}{\Phi(z)} = \sum_{n=-p}^{p} \psi(n)\, z^{-n}.$$

This can be factored by the Fejér method (Robinson [43, p. 194]) as

$$\sum_{n=-p}^{p} \psi(n)\, z^{-n} = \frac{1}{\sigma^2}[1 + \alpha_1 z^{-1} + \cdots + \alpha_p z^{-p}] \cdot [1 + \alpha_1 z + \cdots + \alpha_p z^{p}] \tag{15}$$

where σ^2 is a positive constant and where

$$A(z) = 1 + \alpha_1 z^{-1} + \cdots + \alpha_p z^{-p} \tag{16}$$

is minimum-delay (i.e., where $A(z)$ has no zeros on or outside the unit circle). Thus the maximum-entropy process is specified by

$$\Phi(z) = \frac{\sigma^2}{A(z)\, A(z^{-1})}.$$

This result shows that the maximum entropy process is an AR process of order p.

Silvia and Robinson [83], through the use of lattice methods, have related the concept of maximum entropy to the geophysical inverse problem. Itakura and Saito [84] were responsible for introducing two important ideas into spectrum estimation that are now gaining wide acceptance in the engineering world. The first idea is that of using maximum likelihood in spectrum estimation. Although the idea itself was not new, their introduction of a particular spectrum distance measure is becoming more and more important for different applications, such as speech. Parzen [85] gives the name *information divergence* to this measure, which is the same as the Kullback–Leibler information number. He also shows its relation to the notion of cross-entropy. The second idea is that of using the lattice as a filter structure for the purpose of analysis (as an all-zero filter) and synthesis (as an all-pole filter). The idea of an adaptive lattice was first proposed by Itakura and Saito as a way of estimating the partial correlation (PARCOR) coefficients (a term they coined) adaptively. Makhoul [86] has shown how Burg's technique is really a special case of lattice analysis. Also, the lattice has become important because of its fast convergence and its relative insensitivity to roundoff errors.

XVI. Statistical Theory of Spectrum Estimation

Since the pioneering work of Tukey [34] in 1949, many important contributions have been made to the statistical theory of spectrum estimation. An adequate treatment would require a long paper in itself, and so all we can hope to do here is to raise the reader's consciousness concerning the statistical theory required to understand and implement spectrum estimation.

The writer has great admiration for the work of Parzen, who from the 1950's to the present time has consistently made bedrock contributions both in theory and applications [61], [85], [87], [88]. His long series of papers on time series analysis include the famous Parzen window for spectrum analysis. Another one of Parzen's important contributions is his formulation of the time series analysis problem in terms of reproducing kernel Hilbert spaces. A remarkable number of Ph.D. theses on time series analysis have been written under the direction of Parzen, more than any other person. The writer has had the good fortune to discuss geophysical time-series problems with Professor Parzen over the years, and in every case Parzen has been able to provide important physical insight in the application of the statistical methods. The Harvard lectures by Professor Parzen in 1976 represent one of the high points in time series analysis and spectrum estimation ever to be heard in those venerable halls.

The book by Grenander and Rosenblatt [65] in 1957 formalized many of the data analysis procedures and approximations that have come into use. They have an extensive treatment of the problem of choice of window and bandwidth. The further contributions to this problem by Parzen and by Jenkins are discussed in the 1961 paper of Tukey [51]. An accurate and informative account of the developments of spectrum estimation in the 1950's is given by Tukey [44].

Another important statistical development that deserves mention is the alignment issue in the estimation of coherence, worked on in the 1960's by Akaike and Yamanouch, by Priestley, and by Parzen. Discussions of this work and the references can be found in the book by Priestley [89]. This excellent book, which appeared in 1981, has already set a new standard. It can be recommended as an authoritative account of the statistical theory of spectrum estimation, which we only touch upon in this section.

H. Wold coined the names "moving average process" and "autoregressive process" in his 1938 thesis [90] under Professor Harald Cramér at Stockholm University. In his thesis, Wold computed a model of the yearly level of Lake Vaner in Sweden as a moving average of the current rainfall and the previous year's rainfall. He also computed an autoregressive model of the business cycle in Sweden for the years 1843–1913. In turn, Whittle wrote his 1951 thesis [91] under Professor Wold at Uppsala University. Whittle opened up and made important contributions to the field of hypothesis testing in time-series analysis. Whittle's careful work is exemplified by his autoregressive analysis of a seiche record [92] in which he fits a low level autoregressive model to the data and gives statistical tests to determine the appropriateness of the model. Professor Whittle used to return to Sweden for visits, and the writer remembers taking long walks with him through the Uppsala countryside exploring for old runestones and ancient Viking mounds. Although the writer had left the University of Wisconsin to work with Professor Wold in Sweden, it turned out that Wisconsin under the leadership of Professor G. Box became the real center of time series analysis. It was the joint work of Box with Professor G. M. Jenkins [93] that actually brought the autoregressive (AR) process and the moving average (MA) process to the attention of the general scientific community. The brilliance of this work has made the names Box–Jenkins synonymous with time-series analysis. No achievement is better deserved. No person understands data better than Box in the application of statistical methods to obtain meaningful results.

In the 1960's Parzen [87] and Akaike [94] discussed autoregressive spectrum estimation, and this work led to their crucial work on autoregressive order-determining criteria [88] and [95]. Such criteria have made possible the widespread application of autoregressive spectrum estimation by researchers in diverse scientific fields. Akaike has provided a link between statistics and control theory with deep and significant results, and his work is of the highest tradition that science can provide. Young research workers can learn much by studying his writings well.

We wish we had more space and knowledge to expand upon this section, and those many statisticians whom we have not mentioned should remember that this history is by no means the final word. Someday we hope to write more fully on this subject, and we welcome all comments and suggestions.

XVII. Engineering Use of Spectral Estimation

The purpose of this section is only to refer to the rest of the papers in this special issue of the *Proceedings of the IEEE*. These other papers cover the engineering use of spectral estimation much better than we could do here. There papers represent a living history of the present status of spectral estimation, and, in them and in the references which they give, the reader can find the works of the people who have made spectral analysis and estimation a vital scientific discipline today. As general references, we would especially like to mention the 1978 IEEE book edited by Childers [96], Haykin [97], the *RADC Spectrum Estimation Workshop* [98], the *First IEEE ASSP Workshop on Spectral Estimation* [99], and Ulrych and Bishop [100]. Although much progress has been made, much work yet remains to be done, and there is adventure for a research worker who sets his course in this rewarding and exciting field.

Acknowledgment

I want to express my sincere appreciation to Prof. D. R. Brillinger who let me freely use his paper, "Some history of data analysis of time series in the United States," in *History of Statistics in the United States*, edited by D. B. Owen and published by Marcel Dekker in 1976. I want to thank Dr. J. Makhoul for sending me notes on maximum likelihood and lattice networks. I want to especially thank the authors cited in the Reference Section whose constructive comments materially improved this paper. In writing an historical paper, we should include a thousand references instead of one hundred, so important statistical contributions have unfortunately been left out. Of course, we have purposely not included engineering contributions (as they are covered in the rest of this special issue) but there is never a clear cut line, and so in this sense other important work has also been left out. However, all such omissions are not intentional, and we will gladly try to rectify any situation in some appropriate future publication. Finally, most of all, I want to thank Prof. J. Tukey for the support and help he gave me on spectrum estimation thirty years ago at MIT for which I am forever grateful.

References

[1] I. Newton, *Optics*. London, England, 1704.
[2] B. Taylor, *Methodus Incrementorum Directa et Inverse*. London, England, 1715.
[3] D. Bernoulli, *Hydrodynamics*. Basel, Switzerland, 1738.
[4] L. Euler, *Institutiones Calculi Differentialis*. St. Petersburg, Russia, 1755.
[5] J. L. Lagrange, *Théorie des Fonctions Analytiques*. Paris, France, 1759.
[6] J. Fourier, *Théorie Analytique de la Chaleur*. Paris, France: Didot, 1822.
[7] C. Sturm, "Memoire sur les équations différentielles linéaires du second ordre," *Journal de Mathématiques Pures et Appliquées*, Paris, France, Series 1, vol. 1, pp. 106–186, 1836.
[8] J. Liouville, "Premier mémoire sur la théorie des équations différentielles linéaries et sur le developpement des fonctions en séries," *Journal de Mathématiques Pures et Appliquées*, Paris, France, Series 1, vol. 3, pp. 561–614, 1838.
[9] G. Green, *Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*. Nottingham, England, 1828.
[10] E. Schrödinger, *Collected Papers on Wave Mechanics*. London, England: Blackie, 1928.
[11] W. Heisenberg, *The Physical Principles of Quantum Theory*. Chicago, IL: University of Chicago Press, 1930.
[12] J. von Neumann, "Eigenwerttheorie Hermitescher Funcktionaloperatoren," *Math. Ann.*, vol. 102, p. 49, 1929.
[13] —, *Mathematische Grundlagen der Quantenmechanik*. Berlin, Germany: Springer, 1932.
[14] M. von Smoluchowski, *The Kinetic Theory of Matter and Electricity*. Leipzig and Berlin, Germany, 1914.
[15] A. Einstein, "On the theory of the Brownian movement," *Annalen der Physik*, vol. 19, pp. 371–381, 1906.
[16] N. Wiener, "Differential space," *J. Math. Phys.*, vol. 2, p. 131.
[17] A. Schuster," On the investigation of hidden periodicities with application to a supposed 26-day period of meterological phenomena," *Terr. Magnet.*, vol. 3, pp. 13–41, 1898.
[18] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Phil. Trans. Roy. Soc. London*, A, vol. 226, pp. 267–298.
[19] E. A. Robinson, *Predictive Decomposition of Time Series with Applications to Seismic Exploration*, MIT Geophysical Analysis Group, 1954; Reprinted in *Geophysics*, vol. 32, pp. 418–484, 1967.
[20] —, *An Introduction to Infinitely Many Variates*. London, England: Griffin, 1959, p. 109.
[21] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258, 1930.
[22] A. Y. Khintchine, "Korrelations theorie der Stationären Stochastischen Prozesse," *Math. Ann.*, vol. 109, p. 604.
[23] N. Wiener, *Cybernetics*. Cambridge, MA: MIT Press, 1948.
[24] —, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, MIT NDRC Report, 1942, Reprinted, MIT Press, 1949.

[25] ——, *The Fourier Integral*. London, England: Cambridge, 1933.
[26] P. Dirac, *Principles of Quantum Mechanics*. New York: Oxford University Press, 1930.
[27] O. Heaviside, *Electrical Papers*, vol. I and II. New York: Macmillan, 1892.
[28] J. von Neumann, "Uber Funktionen von Funktional operatoren," *Ann. Math.*, vol. 32, p. 191.
[29] N. Wiener, *I Am a Mathematician*. Cambridge, MA: MIT Press, 1956.
[30] N. Levinson, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *Journal of Math. and Physics*, vol. 26, pp. 110–119, 1947.
[31] N. Levinson, The Wiener RMS (root mean square) error criterion in filter design and prediction, *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.
[32] G. P. Wadsworth, *Short-Range and Extended Forecasting by Statistical Methods*, Air Weather Service, Washington, DC, 1948.
[33] H. R. Seiwell, "The principles of time series analyses applied to ocean wave data," in *Proc. Nat. Acad. Sci., U.S.*, vol. 35, pp. 518–528, 1949.
[34] J. W. Tukey, "The sampling theory of power spectrum estimates," in *Proc. Symp. Appl. Autocorr. Anal. Phys. Prob. U.S. Off. Naval Res.* (NAVEXOS-P-725), 1949. Reprinted in *J. Cycle Res.*, vol. 6, pp. 31–52, 1957.
[35] G. P. Wadsworth, E. A. Robinson, J. G. Bryan, and P. M. Hurley, "Detection of reflections on seismic records by linear operators," *Geophys*, vol. 18, pp. 539–586, 1953.
[36] J. W. Tukey and R. W. Hamming, "Measuring noise color 1," *Bell Lab. Memo*, 1949.
[37] J. W. Tukey, *Measuring Noise Color*, Unpublished manuscript prepared for distribution at the Institute of Radio Engineers Meeting, Nov. 1951.
[38] S. M. Simpson, *Time Series Computations in FORTRAN and FAP*. Reading, MA: Addison-Wesley, 1966.
[39] N. Wiener, *Bull. Amer. Math. Soc.*, vol. 72, pp. 1–145.
[40] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes," *Acta Math.*, vol. 98, pp. 111–150, 1957.
[41] ——, "On bivariate stationary processes," *Theory Prob. Appl.*, vol. 4, pp. 300–308.
[42] R. A. Wiggins and E. A. Robinson, "Recursive solution to the multichannel filtering problem," *J. of Geophys. Res.*, vol. 70, pp. 1885–1891, 1965.
[43] E. A. Robinson, *Statistical Communication and Detection with Special Reference to Digital Data Processing of Radar and Seismic Signals*. London, England: Griffin, 1967. Reprinted with new title, *Physical Applications of Stationary Time Series*. New York: Macmillan, 1980.
[44] J. W. Tukey, "An introduction to the calculations of numerical spectrum analysis," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 25–46.
[45] H. Press and J. W. Tukey, "Power spectral methods of analysis and their application to problems in airplane dynamics," *Bell Syst. Monogr.*, vol. 2606, 1956.
[46] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communications engineering," *Bell Syst. Tech. J.*, vol. 33, pp. 185–282, 485–569, 1958; also New York: Dover, 1959.
[47] J. W. Tukey, "The estimation of power spectra and related quantities," *On Numerical Approximation*, R. E. Langer, Ed. Madison, WI: University of Wisconsin Press, 1959, pp. 389–411.
[48] ——, "An introduction to the measurement of spectra," in *Probability and Statistics*, U. Grenander, Ed. New York: Wiley, 1959, pp. 300–330.
[49] ——, "Equalization and pulse shaping techniques applied to the determination of the initial sense of Rayleigh waves," in *The Need of Fundamental Research in Seismology*, Appendix 9, Department of State, Washington, DC, 1959, pp. 60–129.
[50] B. P. Bogert, M. J. Healy, and J. W. Tukey, "The frequency analysis of time series for echoes; cepstrum pseud-autocovariance, cross-cepstrum and shape-cracking," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, 1963, pp. 201–243.
[51] J. W. Tukey, "Discussion emphasizing the connection between analysis of variance and spectrum analysis," *Technometrics*, vol. 3, pp. 1–29, 1961.
[52] ——, "The future of data analysis," *Ann. Math. Statist.*, vol. 33, pp. 1–67, 1963.
[53] ——, "What can data analysis and statistics offer today?" in *Ocean Wave Spectra*, Nat. Acad. Sci., Washington, DC, and Prentice-Hall, Englewood Cliffs, NJ, 1963.
[54] ——, "Uses of numerical spectrum analysis in geophysics," *Bull Int. Statist. Inst.*, vol. 39, pp. 267–307, 1965.
[55] ——, "Data analysis and the frontiers of geophysics," *Science*, vol. 148, pp. 1283–1289, 1965.
[56] W. J. Pierson and L. J. Tick, "Stationary random processes in meteorology and oceanography," *Bull. Int. Statist. Inst.*, vol. 35, pp. 271–281, 1957.
[57] N. R. Goodman, "On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process," Science Paper no. 10, Engineering Statistics Laboratory, New York University, New York, 1957.
[58] W. H. Munk and G. J. F. MacDonald, *The Rotation of the Earth*. New York: Cambridge University Press, 1960.
[59] M. J. Healy and B. P. Bogert, "FORTRAN subroutines for time series analysis," *Commun. Soc. Computing Machines*, vol. 6, pp. 32–34, 1963.
[60] E. C. Bullard, F. E. Ogelbay, W. H. Munk, and G. R. Miller, *A User's Guide to BOMM*, Institute of Geophysics and Planetary Physics, University of California Press, San Diego, 1966.
[61] E. Parzen, *Time Series Analysis Papers*. San Francisco, CA: Holden-Day, 1967.
[62] E. A. Robinson, *Multichannel Time Series Analysis with Digital Computer Programs*. San Francisco, CA: Holden-Day, 1967.
[63] C. W. J. Granger, *Spectral Analysis of Economic Time Series*. Princeton, NJ: Princeton University Press, 1964.
[64] J. Neyman and E. L. Scott, "Statistical approach to problems of cosmology," *J. Roy. Statist. Soc.*, Series B, vol. 20, pp. 1–43, 1958.
[65] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.
[66] N. Wiener, *Nonlinear Problems in Random Theory*. Cambridge, MA: MIT Press, 1958.
[67] ——, "Rhythm in physiology with particular reference to encephalography," *Proc. Roy. Virchow Med. Soc.*, NY, vol. 16, pp. 109–124, 1957.
[68] J. W. Cooley and J. W. Tukey," An algorithm for the machine calculation of Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, 1965.
[69] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Historical notes on the fast Fourier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 76–79, 1967.
[70] C. Bingham, M. D. Godfrey, and J. W. Tukey, "Modern techniques of power spectrum estimation," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 56–66, 1967.
[71] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE Spectrum*, pp. 63–70, 1967.
[72] *IEEE Trans. Audio Electroacoust.* (Special issue on Fourier transform), B. P. Bogert and F. Van Veen, Eds., vol. AU-15, June 1967 and vol. AU-17, June 1969.
[73] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart, and Winston, Inc., 1974; revised edition by Holden-Day, San Francisco, CA, 1980.
[74] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of k-th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 189–232.
[75] J. Capon, "Applications of detection and estimation theory to large array seismology," *Proc. IEEE*, vol. 58, pp. 760–770, 1970.
[76] *Stochastic Point Processes*, P. A. W. Lewis, Ed. New York: Wiley, 1972.
[77] P. A. W. Lewis, A. M. Katcher, and A. H. Weiss, "SASE IV—An improved program for the statistical analysis of series of events," *IBM Res. Resp.*, RC2365, 1969.
[78] Proc. Symp. Walsh Functions, 1970–1973.
[79] A. Cohen and R. H. Jones, "Regression on a random field," *J. Amer. Statist. Ass.*, vol. 64, pp. 1172–1182, 1969.
[80] J. P. Burg, *Maximum Entropy Spectral Analysis*. Oklahoma City, OK, 1967.
[81] ——, *Maximum Entropy Spectral Analysis*. Stanford, CA: Stanford University, 1975.
[82] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, 1971.
[83] M. T. Silvia and E. A. Robinson, *Deconvolution of Geophysical Time Series in the Exploration of Oil and Natural Gas*. Amsterdam, The Netherlands: Elsevier, 1979.
[84] F. Itakura and S. Saito, "Analysis synthesis telephony based on on the maximum-likelihood method," in *Proc. Sixth Int. Conf. on Acoustics*, (Tokyo, Japan), 1968.
[85] E. Parzen, "Modern empirical spectral analysis," Tech. Report N-12, Texas A&M Research Foundation, College Station, TX, 1980.
[86] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 423–428, 1977.
[87] E. Parzen, "An approach to empirical time series," *J. Res. Nat. Bur. Stand.*, vol. 68B, pp. 937–951, 1964.
[88] ——, "Some recent advances in time series modeling," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 723–730, 1974.
[89] M. Priestley, *Spectral Analysis and Time Series*, 2 volumes, London, England: Academic Press, 1981.
[90] H. Wold, *A Study in the Analysis of Stationary Time Series*. Stockholm, Sweden: Stockholm University, 1938.
[91] P. Whittle, *Hypothesis Testing in Time-Series Analysis*. Uppsala, Sweden: Almqvist and Wiksells, 1951.

[92] —, The statistical analysis of a seiche record, *J. Marine Res.*, vol. 13, pp. 76–100, 1954.
[93] G. Box and G. M. Jenkins, *Time Series Analysis, Forecasting, and Control.* Oakland, CA: Holden-Day, 1970.
[94] H. Akaike, "Power spectrum estimation through autoregressive model fitting," *Ann. Inst. Statist. Math.*, vol. 21, pp. 407–419, 1969.
[95] —, "A new look at statistical model identification," *IEEE Trans. Autom. Contr.*, vol. AC-19, pp. 716–723, 1974.
[96] D. G. Childers, *Modern Spectrum Analysis.* New York: IEEE Press, 1978.
[97] S. Haykin, *Nonlinear Methods of Spectral Analysis.* Berlin, Germany: Springer, 1979.
[98] *Proc. of the RADC Spectrum Estimation Workshop*, Rome Air Development Center, Griffiss Air Force Base, NY, 1979.
[99] *Proc. of the First ASSP Workshop on Spectral Estimation*, McMaster University, Hamilton, Ontario, Canada, 1981.
[100] T. J. Ulrych and T. N. Bishop, "Maximum entropy spectral analysis and autoregressive decomposition," *Rev. Geophys.*, vol. 13, pp. 183–200, 1975.
[101] P. R. Gutowski, E. A. Robinson, and S. Treitel, "Spectral estimation: Fact or fiction," *IEEE Trans. Geosci. Electronics*, vol. GE-16, pp. 80–84, 1978. Reprinted in D. G. Childers, *Modern Spectrum Analysis.* New York: IEEE Press, 1978.

Spectral Estimation: An Overdetermined Rational Model Equation Approach

JAMES A. CADZOW, SENIOR MEMBER, IEEE

Abstract—In seeking rational models of time series, the concept of approximating second-order statistical relationships (i.e., the Yule–Walker equations) is often explicitly or implicitly invoked. The parameters of the hypothesized rational model are typically selected so that these relationships "best represent" a set of autocorrelation lag estimates computed from time series observations. One of the objectives of this paper will be that of establishing this fundamental approach to the generation of rational models.

An examination of many popular contemporary spectral estimation methods reveals that the parameters of a hypothesized rational model are estimated upon using a "minimal" set of Yule–Walker equation evaluations. This results in an undesired parameter hypersensitivity and a subsequent decrease in estimation performance. To counteract this parameter hypersensitivity, the concept of using more than the minimal number of Yule–Walker equation evaluations is herein advocated. It is shown that by taking this overdetermined parametric evaluation approach, a reduction in data-induced model parameter hypersensitivity is obtained, and a corresponding improvement in modeling performance results. Moreover, upon adapting a singular value decomposition representation of an extended-order autocorrelation matrix estimate to this procedure, a desired model order determination method is obtained and a further significant improvement in modeling performance is achieved. This approach makes possible the generation of low-order high-quality rational spectral estimates from short data lengths.

I. Introduction

IN A VARIETY of applications such as found in radar Doppler processing, adaptive filtering, speech processing, underwater acoustics, seismology, econometrics, spectral estimation, and array processing, it is desired to estimate the statistical characteristics of a wide-sense stationary time series. More often than not, this required characterization is embodied in the time series' *autocorrelation lag* sequence as specified by

$$r_x(n) = E\{x(n+m)\,\bar{x}(m)\} \tag{1.1}$$

in which E and $^-$ denote the operations of expectation and complex conjugation, respectively. From this definition, the well-known property that the autocorrelation lags are complex conjugate symmetric (i.e., $r_x(-n) = \bar{r}_x(n)$) is readily established. We will automatically assume this property whenever negative lag autocorrelation elements (or their estimates) are required.

The second-order statistical characterization as represented by the autocorrelation sequence may be given an "equivalent" frequency-domain interpretation. Namely, upon taking the Fourier transform of the autocorrelation sequence, that is,

$$S_x(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_x(n)\, e^{-jn\omega} \tag{1.2}$$

we obtain the associated *power spectral density function* $S_x(e^{j\omega})$ in which the normalized frequency variable ω takes on values in $[-\pi, \pi]$. This function possesses a number of salient properties among which are that it is a positive semi-definite, symmetric (if the time series is real valued), and periodic function of ω. This function is seen to have a Fourier series interpretation in which the autocorrelation lags play the role of the Fourier coefficients. It, therefore, follows that these coefficients may be determined from the power spectral density function through the Fourier series coefficient integral expression

$$r_x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{j\omega})\, e^{j\omega n}\, d\omega. \tag{1.3}$$

Manuscript received May 24, 1982. This work was supported in part by the Signal Processing Section of RADC under Grant AFOSR-81-0250 and by the Statistics and Probability Program of the Office of Naval Research under Contract N00014-82-K-0257.

The author is with the College of Engineering and Applied Sciences, Department of Electrical and Computer Engineering, Arizona State University, Tempe, AZ 85287.

0018-9219/82/0900-0907$00.75