JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 91, 444–509 (1983)

Harmonic Analysis and Spectral Estimation

JOHN J. BENEDETTO*

Department of Mathematics, University of Maryland, College Park, Maryland 20742

Submitted by G.-C. Rota

INTRODUCTION

Spectral estimation problems arise naturally in time series and signal processing, and have long received deep and fruitful attention from statisticians and engineers. We shall introduce some techniques and points of view from harmonic analysis, which have not yet played a major role in spectral estimation but which seem effective for the particular problems we have considered. An example of technique is Wiener's Tauberian theorem; an example of point of view is thinking of spectral estimation in terms of the uniqueness theory associated with the work of Levinson and Beurling.

Our discussion of uniqueness is in Section III; it is the theoretical basis for our new spectral estimation algorithm which we develop in Section II. Spectral estimation involves finding periodicities in noisy signals, and our algorithm, which depends on frequency deconvolution, has motivated the finer deterministic results of Section IV, where we systematically use the Tauberian theorem. Section I contains preliminaries including an outline of generalized harmonic analysis.

The paper is organized as follows:

* This work was supported by the General Research Board of the University of Maryland and by the National Science Foundation.

0022-247X/83/020444-66$03.00/0

For the remainder of this introduction, we shall point out some salient features of the paper as well as some conscious omissions.

(1) The *algorithm* of Section II is a *theorem* in case data $x(t, \alpha)$ are available on the direct product $[-T, T] \times P$, where $[-T, T]$ is a time interval and P is a probability space, e.g., Remark II.8.3(b) and Section III.2.3. In particular, if peaks amidst a broad band spectrum can be detected from the *flatland* $[-T, T] \times P$, then our method will detect them.

(2) The optimistic claim of Item (1) has to be tempered in practice when, for example, only one sample function $x(t)$ may be available on $[-T, T]$, or when it may be necessary to deal with digital simulation or digital input. Such situations breed well known errors in spectral estimation, and the formulation of our algorithm in Section II.8 does not directly treat them. On the other hand, the algorithm does depend on several parameters designed to control these errors. For example, the parameter Γ of Section II.7 can be used to give estimates of the spectrum over long ranges of frequencies, and can also be adjusted to obtain fine resolution on small ranges, e.g., Example II.8.1.

(3) The examples herein are concerned with theoretical matters. Numerical examples of the effectiveness of our algorithm will appear

separately. These examples will include spectra consisting of two close peaks plus white noise as well as the spectra arising in convection and vortex fluids problems using digital data obtained from the numerical analysis of the Navier–Stokes equations, cf. the Acknowledgment below.

(4) Since spectral estimation involves the detection of frequencies using approximants whose relevant frequencies are close to the desired ones, there is a natural kinship between spectral estimation and classical spectral synthesis, e.g., Proposition I.4.2 and Example I.4.1.

(5) Parts of Sections III.9 and IV.8 deal with the Riemann ζ function. We have included these results to indicate some applications to analytic number theory of the methods formulated here, cf. [6], and to give interesting examples of nonperiodic deconvolution in addition to those arising in spectral estimation.

(6) The methods of Section IV, especially Theorem IV.2.1, are used to develop a technique for power spectrum computation in Theorem IV.7.1 and to generalize Wiener's Tauberian theorem, e.g., Remark IV.2.1(b) and the beginning of Section IV.7. Further, the uniqueness theory of Section III not only forms the basis for our algorithm, but combines with Theorem IV.2.1 in Section IV.9 to establish a causal relationship between the future and past behavior of a signal x. This relationship is formulated in terms of the structure of the set of finite frequencies from which the power of x is drawn.

(7) Finally, parts of the paper are expository, e.g., Sections I.3 and III.3. We hope that the exposition puts in perspective some of the diverse topics we have linked together.

I. Preliminaries for Spectral Estimation

I.1. *Notation and Definitions*

We denote by $\mathbb{R}$ and $\hat{\mathbb{R}}$ the real line: $\mathbb{R}$ is the time domain and $\hat{\mathbb{R}}$ is the frequency domain. Further, $L^p(\mathbb{R})$, $1 \leqslant p < \infty$, is the space of Lebesgue measurable functions f on $\mathbb{R}$ for which $\|f\|_p = (\int_{-\infty}^{\infty} |f(t)|^p \, dt)^{1/p} < \infty$, and $M(\hat{\mathbb{R}})$ is the space of bounded Radon measures S on $\hat{\mathbb{R}}$ with total variation norm $\|S\|_1$. The support of an element S from one of these spaces is denoted by supp S. Let $C_0(\hat{\mathbb{R}})$ be the space of continuous functions F on $\hat{\mathbb{R}}$ for which $F(\pm\infty) = 0$. If $C_0(\hat{\mathbb{R}})$ is taken with the sup-norm $\| \;\|_\infty$, then its dual space is $M(\hat{\mathbb{R}})$. In general, we shall use brackets $\langle \;,\; \rangle$ to denote the functional relation between elements of dual spaces; in particular, we write $\langle S, F \rangle$ for $S \in M(\hat{\mathbb{R}})$, $F \in C_0(\hat{\mathbb{R}})$. Also, if $K \subseteq \hat{\mathbb{R}}$ is compact, then we set $\|F\|_{\infty,K} = \sup\{|F(\omega)|: \omega \in K\}$.

If $f \in L^1(\mathbb{R})$, then its Fourier transform $\hat{f} = F$ is

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-it\omega}\, dt,$$

and $A(\hat{\mathbb{R}}) = \{\hat{f} : f \in L^1(\mathbb{R})\}$. The norm of each $\hat{f} \in A(\hat{\mathbb{R}})$ is $\|f\|_1$ and the dual space of $A(\hat{\mathbb{R}})$ is $A'(\hat{\mathbb{R}})$, the space of pseudomeasures with canonical norm $\|\ \|_{A'}$. The convolution $f * g \in L^1(\mathbb{R})$ of $f, g \in L^1(\mathbb{R})$ is $f * g(t) = \int f(u)\, g(t-u)\, du$, and if $\hat{f} = F$ and $\hat{g} = G$, then $(f * g)\hat{} = FG$. If $F \in L^1(\hat{\mathbb{R}})$, then $\|F\|_1 = (1/(2\pi)) \int |F(\omega)|\, d\omega$ and $\hat{F}^{-1}(t) = (1/(2\pi)) \int F(\omega)\, e^{i\omega t}\, d\omega$. In particular, the Lebesgue measure of $[\omega_1, \omega_2] \subseteq \hat{\mathbb{R}}$ is $(1/(2\pi))(\omega_2 - \omega_1)$, and if f, $F \in L^1$ and $\hat{f} = F$, then $\hat{F}^{-1} = f$. Similarly, if $f, g \in L^1(\mathbb{R})$ and $\hat{f} = F$, $\hat{g} = G \in L^1(\hat{\mathbb{R}})$, then $(fg)\hat{}(\omega) = F * G(\omega) = (1/(2\pi)) \int F(\lambda)\, G(\omega - \lambda)\, d\lambda$, e.g., Eq. (II.7.1). Finally, if $S \in M(\hat{\mathbb{R}})$, then its Fourier transform is $\hat{S}(t) = \langle S, e^{-it\omega} \rangle$, cf. (I.2.3).

If f is a locally integrable function, then f' denotes its distributional derivative and $\{f\}'$ is its pointwise derivative, wherever this exists. In the case of locally absolutely continuous functions f, we have $f' = \{f\}'$ distributionally. We let H_γ be the Heaviside function: $H_\gamma(\omega) = 1$ if $\omega \geqslant \gamma$ and $H_\gamma(\omega) = 0$ if $\omega < \gamma$; and we define $H = H_0$ and $H_-(\omega) = H(-\omega)$. Then $H'_\gamma = \delta_\gamma$, the Dirac delta measure at γ. Similarly, δ'_γ is defined as $\langle \delta'_\gamma, F \rangle = -F'(\gamma)$ for a continuously differentiable function F. Finally, we set $\chi_{[\lambda,\gamma)} = H_\lambda - H_\gamma$ and $\chi_T = \chi_{[-T,T)}$.

I.2. *Measure Theory*

Given $S \in M(\hat{\mathbb{R}})$, it is well known that

$$S = s + \sum d_\gamma \delta_\gamma + \mu, \tag{I.2.1}$$

where $s \in L^1(\hat{\mathbb{R}})$, $\sum |d_\gamma| < \infty$, and $\mu \in M(\hat{\mathbb{R}})$ is continuous and singular. The term $\sum d_\gamma \delta_\gamma$, resp. s, is the *discrete*, resp. *absolutely continuous*, part of the complex measure S. Also, there is a function F_S, having bounded variation on $\hat{\mathbb{R}}$, such that

$$F_S = F_s + \sum d_\gamma H_\gamma + F_\mu, \tag{I.2.2}$$

where F_s is absolutely continuous on $\hat{\mathbb{R}}$, F_μ is continuous and has bounded variation on $\hat{\mathbb{R}}$, $F'_S = S$ and $F'_s = s$ with distributional differentiation, and $\{F_\mu\}' = 0$ a.e. If $G \in C_0(\hat{\mathbb{R}})$, then the duality $\langle S, G \rangle$ has the representation

$$\begin{aligned} \langle S, G \rangle &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega)\, dF_S(\omega) \qquad\qquad (I.2.3) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} s(\omega)\, G(\omega)\, d\omega + \frac{1}{2\pi} \sum d_\gamma G(\gamma) + \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega)\, dF_\mu(\omega), \end{aligned}$$

where the first and last integrals are Riemann–Stieltjes integrals and $\int s(\omega)\,G(\omega)\,d\omega$ is the Lebesgue integral. We sometimes write $\langle S, G\rangle = (1/2\pi)\int G\,dS$.

Translation of S by γ is defined as $\langle \tau_\gamma S, G\rangle = \langle S, \tau_{-\gamma} G\rangle$, where $\tau_{-\gamma}G(\omega) = G(\omega+\gamma)$; and the convolution $S * G(\lambda)$ is

$$S * G(\lambda) = \langle S_\omega, G(\lambda-\omega)\rangle = (1/2\pi)\int_{-\infty}^{\infty} G(\lambda-\omega)\,dF_S(\omega),$$

where S_ω indicates the dependence of S on ω. More generally, if $S_1, S_2 \in M(\hat{\mathbb{R}})$, then the convolution $S_1 * S_2 \in M(\hat{\mathbb{R}})$ is

$$\langle S_1 * S_2, G\rangle = \langle (S_1)_\lambda, \langle (S_2)_\omega, G(\lambda+\omega)\rangle\rangle;$$

and so $\delta_\gamma * S = \tau_\gamma S$.

EXAMPLE I.2.1. Take $G \in C_0(\hat{\mathbb{R}})$ and S as in (I.2.1). Then $S * (GH)$ is well defined. In fact, we have

$$\begin{aligned} S * (GH)(\lambda) &= s * (GH)(\lambda) + \frac{1}{2\pi}\sum d_\gamma (GH)(\lambda-\gamma) \\ &\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} (GH)(\lambda-\omega)\,dF_\mu(\omega) \\ &= \frac{1}{2\pi}\int_0^{\infty} s(\lambda-\omega)\,G(\omega)\,d\omega \\ &\quad + \frac{1}{2\pi}\sum_{\gamma\leqslant\lambda} d_\gamma G(\lambda-\gamma) + \frac{1}{2\pi}\int_{-\infty}^{\lambda} G(\lambda-\omega)\,dF_\mu(\omega), \end{aligned}$$

where all of the integrals clearly exist.

A measure space $(P, \mathscr{P}, p)$ for which $p(P) = 1$ is a *probability space*; and a function $x\colon \mathbb{R}\times P \to \mathbb{C}$ is a *stochastic process* if $x(t,\alpha)$ is a measurable function on P for each $t\in\mathbb{R}$. The *expected value* of the process x is $E\{x(t)\} = m(t)$, $t\in\mathbb{R}$, where the mean value $E\{x(t)\}$ is defined as $\int_P x(t,\alpha)\,dp(\alpha)$. Suppose $x(t,\cdot)\in L^2(P)$ for each $t\in\mathbb{R}$. Then x is a *stationary stochastic process* (SSP) if $m(t)$ is a constant m and if

$$\forall t, u, h \in \mathbb{R}, \qquad E\{x(t+h)\,\overline{x(u+h)}\} = E\{x(t)\,\overline{x(u)}\}.$$

We shall deal exclusively with SSPs x equipped with the continuity property, $\lim_{t\to 0} E\{|x(t)-x(0)|^2\} = 0$.

The *autocorrelation* of the SSP x is the continuous function

$$R(t) = E\{x(u+t)\,\overline{x(u)}\}$$

and the *autocovariance* of x is $C(t) = E\{(x(u+t) - m)\overline{(x(u) - m)}\} = R(t) - |m|^2$. The *variance* of each random variable $x(t, \cdot)$ is $\sigma^2 = C(0) = R(0) - |m|^2$.

Basic texts elaborating the measure theory and probability theory herein are [5 and 27], respectively.

I.3. *Generalized Harmonic Analysis*

Let $L^\infty(\mathbb{R})$ be the space of essentially bounded Lebesgue measurable functions on $\mathbb{R}$. For $x \in L^\infty(\mathbb{R})$, we define the means $p_T(t)$, $T > 0$, as

$$p_T(t) = \frac{1}{2T} \int_{-T}^{T} \overline{x(u)}\, x(t+u)\, du; \tag{I.3.1}$$

and we let $\{P_\alpha\}$ denote the set of $\sigma(L^\infty, L^1)$, i.e., weak $*$, limit points of $\{p_T\colon T > 0\}$. If $\{p_T\}$ has a (single) weak $*$ limit, we denote this limit by P. We shall see that in the case of certain SSPs x, we can assert that P is the autocorrelation R. Because of this connection, we refer to elements $x \in L^\infty(\mathbb{R})$ as *sample functions.*

Given $P \in L^\infty(\mathbb{R})$; P is a function of *positive type* if

$$\forall f \in L^1(\mathbb{R}), \qquad \iint P(t+u) f(t) \tilde{f}(u)\, dt\, du \geqslant 0, \tag{I.3.2}$$

where $\tilde{f}(t) = \overline{f(-t)}$. If P is a pointwise function defined everywhere on $\mathbb{R}$, then P is *positive definite* if

$$\forall t_1, \ldots, t_n \in \mathbb{R} \quad \text{and} \quad \forall c_1, \ldots, c_n \in \mathbb{C}, \qquad \sum c_j \bar{c}_k P(t_j - t_k) \geqslant 0.$$

EXAMPLE I.3.1. If x is an SSP, then the autocorrelation R is clearly a continuous positive definite function. If x does not have the continuity property we have assumed, then R is still positive definite, and it has the decomposition $R = R_C + R_0$, where R_C and R_0 are positive definite, R_C is continuous, and $R_0 = 0$ a.e. This decomposition is due mostly to F. Riesz (*Acta Sci. Math.*, 1933).

The following Lemma gives more examples:

LEMMA I.3.1. (a) *Let P be a continuous function on $\mathbb{R}$. Then P is positive definite, and therefore bounded, if and only if P is a function of positive type.*

(b) *For a given sample function $x \in L^\infty(\mathbb{R})$, each weak $*$ limit point P_α is a function of positive type.*

Proof. (a) The result is easily checked using discrete measures $\sum d_j \delta_{t_j}$ and approximating compactly supported continuous functions f (considered as measures) in the weak $*$ topology on $M(\mathbb{R})$. It is for this step that we use the continuity of P.

(b) For each α, let $\{T_n\}$ have the property that $\lim_n p_{T_n} = P_\alpha$ in the $\sigma(L^\infty, L^1)$ topology. Letting χ_n be the characteristic function of $[-T_n, T_n]$, a direct calculation shows that $\lim_n (1/(2T_n))(\chi_n x) * (\chi_n x)^\sim = P_\alpha$ in the $\sigma(L^\infty, L^1)$ topology. Since $(1/(2T_n))(\chi_n x) * (\chi_n x)^\sim \equiv p_n$ is easily seen to be a continuous positive definite function, we have from part (a) that p_n is a function of positive type. The definition of positive type and the weak $*$ convergence, $\lim p_n = P_\alpha$, yield the conclusion that P_α is a function of positive type. Q.E.D.

THEOREM I.3.1. *Given a sample function $x \in L^\infty(\mathbb{R})$, then, in the notation of* (I.3.1), *for each α there is a positive bounded measure $\mu_\alpha \in M(\hat{\mathbb{R}})$ such that $\hat{\mu}_\alpha = P_\alpha$ a.e.; in particular, there is a continuous function which equals P_α a.e.*

Proof. (i) Because of Lemma I.3.1, each $P_\alpha \in L^\infty(\mathbb{R})$ is a function of positive type. The Bochner–Schwartz theorem [37, pp. 276–277] shows that there is a positive tempered measure μ_α (possibly unbounded at this point) for which $\hat{\mu}_\alpha = P_\alpha$, as distributions. Since $P_\alpha \in L^\infty(\mathbb{R})$, we know that $\mu_\alpha \in A'(\hat{\mathbb{R}})$ and $\hat{\mu}_\alpha = P_\alpha$ a.e.

(ii) We now show that positive pseudomeasures μ_α are elements of $M(\hat{\mathbb{R}})$. We already know that μ_α is a measure (although this also follows easily from the hypothesis that it is a positive pseudomeasure).

The Fejér kernel $\{\hat{f}_n \geqslant 0 : n = 1, \ldots\}$ has the following properties: for each n, $\hat{f}_n \geqslant 0$, $\|f_n\|_1 = 1$, $\operatorname{supp} \hat{f}_n$ is compact, and

$$\forall \gamma \in \hat{\mathbb{R}}, \qquad \lim_{n \to \infty} \hat{f}_n(\gamma) = 1, \tag{I.3.3}$$

cf., Section II.7.

Since $\mu_\alpha \in A'(\hat{\mathbb{R}})$, $\|\mu_\alpha\|_{A'} = \sup\{|\langle \mu_\alpha, \hat{f} \rangle| : f \in L^1(\mathbb{R}),\ \|f\|_1 \leqslant 1$, and $\operatorname{supp} \hat{f}$ is compact$\}$, and $\langle \mu_\alpha, \hat{f} \rangle = \int \hat{f}\, d\mu_\alpha$, we conclude from (I.3.3) and the monotone convergence theorem that $\int d\mu_\alpha \leqslant \|\mu_\alpha\|_{A'} < \infty$. Thus $\mu_\alpha \in M(\hat{\mathbb{R}})$.

(iii) Parts (i) and (ii) combine to complete the proof. Q.E.D.

Remark I.3.1. (a) The *proof* of Theorem I.3.1 is more an outline than not, but we have supplied relevant references. A pointwise version of Theorem I.3.1 was proved by Wiener [43, Theorem 35, p. 183] by mid-1932 and was later proved by F. Riesz (*Acta Sci. Math.*, received June 16, 1933) who refers to Bochner and Stone but not Wiener. The proofs are valid for any positive definite function and do not depend on the setup of (I.3.1). Also, there are two quite general procedures which yield Theorem I.3.1 as a corollary: the Gelfand–Raikov theorem (as formulated in Godement's University of Paris lecture notes) and Varopoulos' theorem on the continuity of positive functionals.

(b) It is not difficult to show that $\{p_T\}$ does not have a weak $*$ limit for the case $x(t) = \chi_{(0,\infty)}(t)\, e^{it \log t}$. In general, the supports $\operatorname{supp} \mu_\alpha$ vary for each α; this fact was proved by Bertrandias (1973) and pointed out to the author in the present context by Ka-Sing Lau.

If $x \in L^\infty(\mathbb{R})$ is a sample function for which $\{p_T\}$ has a weak $*$ limit P or if x is an SSP with autocorrelation R, then we denote the positive measure μ, established in Theorem I.3.1, by S; thus, $\hat{S} = P$ a.e. or $\hat{S} = R$, respectively. In either case, we say that S is the *power spectrum* of x and $W = \operatorname{supp} S$ is the *Wiener spectrum* of x.

EXAMPLE I.3.2. (a) Since positive definite functions P satisfy the condition $P(t) = \overline{P(-t)}$, we see that if x is a real sample function or real SSP, then S is an *even measure* in the sense that $\langle S_\omega, F(\omega)\rangle = \langle S, F(-\omega)\rangle$ for all $F \in C_0(\hat{\mathbb{R}})$.

(b) Let $x \in L^\infty(\mathbb{R})$ be a sample function which has a power spectrum S. Then several applications of Fubini's theorem yield

$$\iint P(u-t) f(t) \overline{f(u)}\, dt\, du = \int |\hat{f}(-\gamma)|^2\, dS(\gamma)$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f * x(t)|^2\, dt$$

$$= \int \overline{P(t)}\, f * \tilde{f}(t)\, dt \qquad \text{(I.3.4)}$$

for each $f \in L^1(\mathbb{R})$, where $\hat{S} = P$. Using Wiener's Tauberian theorem, we can verify the middle equality of (I.3.4) for any $f \in M(\mathbb{R})$ for which $\hat{f} \in C_0(\hat{\mathbb{R}})$, e.g., [4, pp. 90ff.].

(c) If $x \in L^\infty(\mathbb{R})$ is a sample function, then $\hat{x} = X \in A'(\hat{\mathbb{R}})$. The analogous result for an SSP x is the spectral representation theorem due to Cramér, Kolmogorov, Karhunen, *et al.*, e.g., [27, pp. 46–47; 30, pp. 86–87]. Bochner's theorem played a role in the original formulation of this result, but this gave way to Hilbert space methods when Karhunen (1947) and Kolmogorov showed that there is a canonical one-parameter unitary group associated with each SSP, thus allowing the application of Stone's extension of the spectral theorem.

An SSP x is a *correlation ergodic process* if

$$\forall t \in \mathbb{R}, \qquad \exists A_t \in \mathscr{P} \qquad \text{such that} \quad p(A_t) = 1$$

and

$$\forall \alpha \in A_t, \qquad \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t+u,\alpha)\,\overline{x(u,\alpha)}\,du = R(t). \qquad \text{(I.3.5)}$$

A proof of the next result, as well as an excellent accompanying exposition, can be found in [34, pp. 352–360].

THEOREM I.3.2. *Let x be an* SSP *with autocorrelation R. Then* (I.3.5) *is satisfied* $\Leftrightarrow$

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-2T}^{2T} C(t,v)\left(1-\frac{|v|}{2T}\right)dv = 0, \qquad \text{(I.3.6)}$$

where $C(t,v) = E\{x(t+u+v)\,\overline{x(u+v)}\,x(t+u)\,\overline{x(u)}\} - |R(t)|^2$.

Remark I.3.2. We first note the appearance of the Fejér kernel again in (I.3.6), cf. the proof of Theorem I.3.1. Also, in the case of a Gaussian process x for which $E\{x(t)\} = 0$, (I.3.6) can be simplified to the condition

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-2T}^{2T} (R(t+v)\,R(t-v) + R(v)^2)\left(1-\frac{|v|}{2T}\right)dv = 0. \qquad \text{(I.3.7)}$$

Finally, if x is an SSP with autocorrelation R and if t is fixed, then

$$\forall T > 0, \qquad \int_P \left(\frac{1}{2T}\int_{-T}^{T} x(t+u,\alpha)\,\overline{x(u,\alpha)}\,du - R(t)\right)dp(\alpha) = 0 \qquad \text{(I.3.8)}$$

since $E\{(1/(2T))\int_{-T}^{T} x(t+u,\cdot)\,x(u,\cdot)\,du\} = R(t)$. Equation (I.3.8) and F. Riesz's result relating convergence in measure to pointwise almost everywhere convergence of subsequences provide tantalizing motivation for establishing Theorem I.3.2.

Suppose x is an SSP which can be written in the form $x(t,\alpha) = f(y(t), n(t,\alpha))$, where $y \in L^\infty(\mathbb{R})$ and n is a stochastic process; x is a *noisy signal* or, more accurately, a (true) signal y with noise n. Frequently, the noise is additive: $x(t,\alpha) = y(t) + n(t,\alpha)$. We close this section by discussing a specific important noise.

EXAMPLE I.3.3. (a) *White noise* $w(t,\alpha)$ is an SSP for which $S(\omega) = c > 0$ for all $\omega \in \hat{\mathbb{R}}$ and $R = \hat{S} = 2\pi c\delta$. The nonstationary case is also important, e.g., [34, pp. 303–321]. Of course, this means that white noise has infinite power ($R(0) = \infty$), a possibility that can cause some apprehension in the real world. In practice, an analog signal is said to have a white spectrum S if S is essentially constant over a long range of frequencies including those of interest in the problem at hand; thus, S could have the form χ_Ω or $\chi_\Omega * f$ for some approximate identity f. Moderately white spectra are often referred to as *pastel* or *pink*.

In the case of a digital signal, precise white noise $w(n, \alpha)$ is possible. For example, one need only take an SSP $w(n, \alpha)$ for which $\{w(n, \alpha): n = 0, \pm 1, ...\}$ is a sequence of independent, identically distributed random variables and each $E\{x(n)\} = 0$. Then $R(n) = E\{w(n+m)\} E\{\overline{w(m)}\} = 0$ if $n \neq 0$ and $R(0) = E\{|x(m)|^2\} = 2\pi c$ independent of m for $n = 0$.

It is frequently possible to reduce computational noise or leakage of power if we transform a given signal so as to flatten out its spectrum; this is the *prewhitening transformation*, e.g., [39, pp. 306–308; 40, pp. 29–30], cf. Remark II.8.2 of Section II and Brillinger [13, p. 512], who compares high resolution methods with classical ones which use prewhitening techniques.

(b) Even though $2\pi c\delta \notin L^\infty(\mathbb{R})$, it is a *measure of positive type* for $c > 0$ since

$$\langle\langle\delta(t+u), f(t)\rangle, \tilde{f}(u)\rangle = \langle\overline{f(-u)}, \langle\delta(t), \tau_u f(t)\rangle\rangle = \langle\overline{f(-u)}, f(-u)\rangle$$

for each $f \in L^2(\mathbb{R}) \cap C_0(\mathbb{R})$. Further, $S(\omega) = c$ on $\hat{\mathbb{R}}$ is a slowly increasing unbounded positive measure, cf. [37, pp. 275–276].

(c) The existence, mathematically, of white noise goes back to Wiener's research on the Wiener or Brownian motion process; and it can be effected in the following way: Let $\mathscr{I}(\mathbb{R})$ be the family of bounded intervals $I \subset \mathbb{R}$. Gaussian processes $W(I, \alpha)$ on $\mathscr{I}(\mathbb{R}) \times P$, for which $E\{W(I)\, W(J)\}$ is the length of $I \cap J$, are known to exist, e.g., [18, pp. 62–63]. Stationary white noise is the formal derivative, $w(t, \alpha) = \lim_n W((a_n, b_n), \alpha)/(b_n - a_n)$, where $t \in (a_n, b_n)$ and $\lim_n(b_n - a_n) = 0$. Classically, this derivative was interpreted by means of analytically sound computations with its primitive, e.g., [17, p. 71; 30, pp. 84–86; 25, p. 90 by Brillinger]. Nowadays, white noise is frequently defined in terms of Hilbert space valued functions on $\mathbb{R}$, e.g., [18, pp. 92ff.; 25, pp. 4ff. by Balakrishnan].

Further developments in generalized harmonic analysis can be found in [30; 4, Chap. 2; 29, pp. 638–640], and by Masani in his essay in Volume 2 of Wiener's *Oeuvres*.

I.4. *Periodograms*

We shall use periodograms in our spectral estimation algorithm. Schuster initiated periodogram analysis and his work was one of the major influences on Wiener's generalized harmonic analysis, e.g., [42, Introduction and Sect. 2]. The periodogram is a natural device for spectral estimation, but in its pristine state it has been subjected to some harsh criticism in this role, e.g., "misleading" and "dangerous" by Tukey [40, pp. 25–26], cf.

Proposition I.4.2 and Example I.4.1. Bartlett and Tukey were responsible for the early adaptation of periodograms as an effective tool in spectral estimation, e.g., Example I.3.3(a) and Remark II.8.2.

Given $b \in L^1(\mathbb{R})$, suppose x is an SSP such that each sample function $x(\cdot, \alpha)$ is an element of $L^\infty(\mathbb{R})$. Then

$$S_b(\omega, \alpha) = \left| \int b(t)\, x(t, \alpha)\, e^{-i\omega t}\, dt \right|^2 \tag{I.4.1}$$

is the *periodogram* associated with the process x and *data window* b. The basic formula which we exploit in Section II is given in

PROPOSITION I.4.1. *Let x be an SSP and suppose $\hat{b} = B \in A(\mathbb{R})$. Then we have*

$$\int_{-\infty}^{\infty} |b(t)|\, \|x(t, \cdot)\|_{L^2(P)}\, dt < \infty \tag{I.4.2}$$

and

$$E\{S_b(\omega)\} = \int |B(\omega + \gamma)|^2\, dS(\gamma). \tag{I.4.3}$$

If x is real, then the right-hand side of (I.4.3) *is* $\int |B(\omega - \gamma)|^2\, dS(\gamma)$; *and if x is real and $b = \tilde{b}$, then*

$$(1/2\pi)\, E\{S_b(\omega)\} = S * B^2(\omega). \tag{I.4.4}$$

Proof. The hypotheses, $b \in L^1(\mathbb{R})$ and x stationary, give (I.4.2), which allows us to use Fubini's theorem several times in the following calculation, which yields (I.4.3):

$$E\{S_b(\omega)\} = \iint b(t)\, \overline{b(u)}\, e^{-i\omega(t-u)} \left(\int e^{-i\gamma(t-u)}\, dS(\gamma) \right) du\, dt$$

$$= \int \left(\iint b(t)\, \overline{b(u)}\, e^{-it(\omega+\gamma)} e^{iu(\omega+\gamma)}\, du\, dt \right) dS(\gamma). \qquad \text{Q.E.D.}$$

Remark I.4.1. (a) In our algorithm, we shall normalize $b = \tilde{b}$ in Proposition I.4.1 so that $(1/(2\pi)) \int B^2(\omega)\, d\omega = 1$. Classically, this is done so that B^2 can resemble an approximate identity and hence $E\{S_b(\omega)\}$ is an estimate, albeit biased, of S; a typical example, used below in Proposition I.4.2, is $b = (1/\sqrt{2T})\, \chi_T$. The data windows we use are quite different from the classical situation, since our purpose in choosing them is to deconvolve effectively, e.g. Section II.7 and Remark II.8.3(b).

(b) Because of (I.4.4) and the fact that in some spectral estimation problems $E\{S_b\}$ is known and an estimate of S is desired, it is tempting to apply the operational calculus to (I.4.4) and solve for S in terms of B^2 and $E\{S_b\}$ using Fourier and inverse Fourier transforms. The problem is that $\operatorname{supp} b * b$ is compact in such problems, so that we would be faced with division by zero. Consequently, we have implemented our frequency deconvolution method in Section II. The operational calculus is global when it works, but when it does not work, as in the case that $\operatorname{supp} b * b$ is compact, the frequency deconvolution method can be used to estimate S locally.

The proof of the following result is found in [4, p. 87]:

PROPOSITION I.4.2. *Let x be a real* SSP *and suppose $b_T = (1/\sqrt{2T})\chi_T$. Then we have*

$$\lim_{T\to\infty} (1/2\pi)\, E\{S_{b_T}(\omega)\} = S \tag{I.4.5}$$

in the weak $*$ topology $\sigma(A', A)$.

EXAMPLE I.4.1. The weak $*$ convergence of (I.4.5) allows for a great deal of mischief on the part of the raw periodogram S_{b_T} if one thinks of it as an approximant to S. There are results due to Beurling, Herz, and Pollard which are similar to (I.4.5), but for which the convergence takes place in stronger topologies. These theorems are useful in spectral synthesis, but it is not yet clear how accurate an approximation they provide to S for a given finite time T. An example of such a theorem for closed intervals I, disjoint from $\operatorname{supp} S$, is

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} R(t)\, e^{it\omega} e^{|t|/T}\, dt = 0, \qquad \text{uniformly on} \quad I. \tag{I.4.6}$$

There are similar statements for other $L^p(\hat{\mathbb{R}})$ norms and other windows, and Herz (1957) even has such a result for the $A(\hat{\mathbb{R}})$ norm in the case of a class of windows including Riemann's $(R, 2)$ summability, cf. [4, esp. Sect. 2.1] for other references and results.

In Section II.1 we shall state precisely the spectral estimation problem for which we provide an algorithm. We close this section by giving a general informal formulation of such problems, of which the statement in Section II.1 is one of many possibilities, and by listing some references which provide a good history of the subject; our list is not unique.

Given some information about a noisy signal $x(t, \alpha)$, the problem of *spectral estimation* is to find a useful estimate of its power spectrum S. In general, the given *information* includes values of x or of a sample function of

x or of R on some set $X \subseteq \mathbb{R}$. The criterion that the estimate be *useful* includes being able to find the peaks of S, i.e., the periodicities in x, cf. the first few pages of [11]. Sometimes observed spectral peaks really represent x, *but* sometimes they arise (sic) because of the estimation procedure.

The work by Blackman and Tukey [11] (1958) is a classic for history, technique, and important insights into spectral estimation; and [40] is a beautiful brief update. This material is essentially periodogram analysis. The high resolution techniques are MEM (maximum entropy method) and MLM (maximum likelihood method). The one-dimensional maximum entropy estimator is equivalent to the autoregressive spectral estimator, e.g., [1]; Burg and Parzen made important contributions on this method in the late 1960s. Lacoss [26] and Baggeroer [2] give fundamental examples and comparisons of MEM and MLM; the book edited by Childers [16] is a basic collection of papers on these methods. Brillinger [13, pp. 511–512] briefly comments on the comparability of the above methods.

II. A Spectral Estimation Algorithm

II.1. *The Problem and Frequency Deconvolution*

For a fixed $T > 0$ and data window $b \in L^2(\mathbb{R})$ for which $b = \tilde{b}$ and $\operatorname{supp} b \subseteq [-T, T]$, we consider the following setup for a given real noisy signal $x(t, \alpha)$ defined on $[-T, T] \times P$. Take the periodogram,

$$S_b(\omega, \alpha) = \left| \int_{-T}^{T} b(t)\, x(t, \alpha)\, e^{-i\omega t}\, dt \right|^2, \tag{II.1.1}$$

so that from Proposition I.4.1 we have

$$(1/2\pi)\, E\{S_b(\omega)\} = S * B^2(\omega) \equiv E_b(\omega), \tag{II.1.2}$$

an infinitely differentiable function, where $\hat{b} = B$. Besides the condition,

$$\operatorname{supp} b \subseteq [-T, T], \tag{II.1.3}$$

we also assume that

$$(1/2\pi) \int_{-\infty}^{\infty} B^2(\omega)\, d\omega = 1 \qquad \text{and} \qquad B(0) \neq 0. \tag{II.1.4}$$

In particular, for a given T and b, E_b *is a computable function and* S *is not generally known.*

Problem. For a given metric ρ on $M(\hat{\mathbb{R}})$ and a given $\varepsilon > 0$, find T and b satisfying conditions (II.1.1)–(II.1.4) and construct a computable function

$K \in C^{\infty}(\mathbb{R}) \cap M(\mathbb{R})$ depending on T and b so that $\rho(S, K) < \varepsilon$. Such a K is a *spectral estimator.*

In practice, we are interested in finding a function K which reflects the discrete part of the measure S or, equivalently, any periodic signal $y(t)$ embedded in the noisy signal $x(t, \alpha)$. This is accomplished in the algorithm we present.

Remark II.1.1. A key feature of the Problem as stated is that T is a *variable* which is to be adjusted to gather information about the *unique* power spectrum which would be obtained if we let the signal $x(t, \alpha)$ run its course for all time. Another point of view, to which autoregressive and high resolution algorithms address themselves, is to consider T *fixed* and to extend x or R beyond $[-T, T]$ in certain intelligent ways so as to find functions whose peaks represent periods within the given noisy signal.

The essential feature of our algorithm is to deconvolve B^2 in (II.1.2). As such, B^2 must be chosen in a special way. The following example indicates a necessary requirement:

EXAMPLE II.1.1. Let $B^2 = B^2(0)\, \chi_{[0,\Gamma)}$ and suppose the convolution,

$$S * \chi_{[0,\Gamma)} = (1/B^2(0))\, E_b,$$

is given data. The inverse of $\chi_{[0,\Gamma)}$ under convolution is

$$\chi_{[0,\Gamma)}^{-1} = \sum_{n=0}^{\infty} \delta'_{n\Gamma};$$

and hence

$$S = \frac{1}{B^2(0)} \sum_{n=0}^{\infty} \delta'_{n\Gamma} * E_b.$$

If $\operatorname{supp} S \subseteq [-\Omega, \Omega]$, then $\operatorname{supp}(E'_b + \tau_\Gamma E'_b + \cdots + \tau_{n\Gamma} E'_b) \subseteq [-\Omega, \Omega + (n+1)\Gamma]$ and $\operatorname{supp} \tau_{n\Gamma} E'_b \subseteq [-\Omega + n\Gamma, \Omega + (n+1)\Gamma]$ for $n \geqslant 0$. Therefore, if $\Gamma > 2\Omega$, then

$$S = (1/B^2(0))\, E'_b \qquad \text{on} \quad (-\Omega, \Omega).$$

Remark II.1.2. (a) Usual periodogram technique uses *even* functions B^2, but this has the effect of making plateaulike convolutions. By choosing B^2 as in Example II.1.1 the convolution $S * B^2$ becomes an increasing function with sharpest increases at peaks of S.

(b) Since we *must* have (II.1.3) we cannot have $\operatorname{supp} B \subseteq [0, \infty)$ as in Example II.1.1. This follows from the second theorem of F. and M. Riesz

(1916): if $f \in L^1(\mathbb{R})$, $\hat{f} = F = 0$ on $(-\infty, \omega_0]$, and the Lebesgue measure of $\{t: f(t) = 0\}$ is positive, then $f = 0$ a.e., e.g., [7; 23, Chap. 4, esp. pp. 50–52; cf. 27, pp. 78–81 for the original proof].

We shall circumvent the potential problems in Remark II.1.2 and construct a spectral estimator K along the lines suggested by Example II.1.1.

II.2. *Deconvolution of Step Functions—Periodic Frequency Sampling*

Given a function $C: \mathbb{R} \to \mathbb{C}$ for which $C(0) = 1$, C vanishes on $(-\infty, 0)$, $C(0+) = 1$, and C is continuous on $(0, \infty)$, fix any $c > 0$ and define the numbers $\alpha_0 = 1$,

$$\beta_j = C(cj), \qquad j \geqslant 0, \tag{II.2.1}$$

and

$$\alpha_n = 1 - \sum_{m=0}^{n-1} \beta_{n-m} \alpha_m, \qquad n \geqslant 1. \tag{II.2.2}$$

Clearly, the α's and the β's depend on c, and the frequency samples $\{jc: j \geqslant 0\}$ are periodic.

Next we define the functions

$$C_{c,n} = \sum_{j=0}^{n} \beta_j \chi_{[jc,(j+1)c)}, \qquad n \geqslant 1, \tag{II.2.3}$$

and the distributions

$$C_{c,n}^{-1} = \sum_{j=0}^{n} \alpha_j \delta'_{jc}, \qquad n \geqslant 1, \tag{II.2.4}$$

on $\mathbb{R}$. Also, we write

$$C_c = C_{c,\infty} \qquad \text{and} \qquad C_c^{-1} = C_{c,\infty}^{-1}. \tag{II.2.5}$$

PROPOSITION II.2.1. (a) $C_c * C_c^{-1} = \delta$.

(b) *For each* $n \geqslant 1$,

$$C_{c,n} * C_{c,n}^{-1} = \delta - \delta_{cn} + \sum_{n \leqslant m \leqslant 2n} \sum_{\substack{j+k=m \\ 0 \leqslant j,k \leqslant n}} \beta_j \alpha_k (\delta_{(j+k)c} - \delta_{(j+k+1)c}).$$

Proof. By definition we compute

$$\begin{aligned} C_{c,n} * C_{c,n}^{-1} &= \sum_{j=0}^{n} \sum_{k=0}^{n} \beta_j \alpha_k (\delta_{(j+k)c} - \delta_{(j+k+1)c}) \\ &= \beta_0 \alpha_0 (\delta_0 - \delta_c) + \sum_{\substack{j+k \geqslant 1 \\ j,k \leqslant n}} \beta_j \alpha_k (\delta_{(j+k)c} - \delta_{(j+k+1)c}). \end{aligned} \tag{II.2.6}$$

Note that $\alpha_0 = \beta_0 = 1$, $\alpha_1 = 1 - \beta_1 \alpha_0 = 1 - C(c)$, and $\alpha_2 = 1 - C(c) - C(2c) + C(c)^2$. Thus, (II.2.6) becomes

$$\delta_0 - \delta_c + (\delta_c - \delta_{2c})(\beta_0 \alpha_1 + \beta_1 \alpha_0) + (\delta_{2c} - \delta_{3c})(\beta_0 \alpha_2 + \beta_1 \alpha_1 + \beta_2 \alpha_0) + \sum_{\substack{j+k \geqslant 3 \\ j,k \leqslant n}}$$

$$\begin{aligned} &= \delta_0 - \delta_c + (\delta_c - \delta_{2c})(1 - C(c) + C(c)) \\ &\quad + (\delta_{2c} - \delta_{3c})([1 - C(c) - C(2c) + C(c)^2] \\ &\quad + [C(c) - C(c)^2] + [C(2c)]) + \sum_{\substack{j+k \geqslant 3 \\ j,k \leqslant n}} = \delta_0 - \delta_{3c} + \sum_{\substack{j+k \geqslant 3 \\ j,k \leqslant n}} = \cdots \\ &= \delta - \delta_{nc} + \sum_{\substack{j+k > n \\ j,k \leqslant n}} \beta_j \alpha_k (\delta_{(j+k)c} - \delta_{(j+k+1)c}), \end{aligned} \tag{II.2.7}$$

and this yields part (b) of the result.

Since the support of the measure $C_{c,n} * C_{c,n}^{-1} - \delta$ (in part (b)) is contained in the interval $[(n+1)c, \infty)$, we see that $C_c * C_c^{-1} - \delta = 0$, and this is part (a). Q.E.D.

The following is clear:

Proposition II.2.2. *If* $C(+\infty) = 0$, *then* $\lim_{c \to 0} C_c = C$ *uniformly on* $\mathbb{R}$.

Example II.2.1. (a) Equation (II.2.2) is the matrix equation $(\beta_{ij})(\alpha_j) = (1)$, where $i, j = 0, 1, \ldots$, $\alpha_0 = \beta_0 = 1$, $\beta_{i,j} = \beta_{i-j}$ for $i \geqslant j$, and $\beta_{i,j} = 0$ for $i < j$.

(b) Setting $B = (\beta_{ij})$ and letting B^t denote the transpose of B we define $C = B + (B^t - I)$, where I is the identity matrix. If the sequence $\{\beta_j\}$ were autocovariance data, then C would be the Yule–Walker matrix which plays an important role in autoregressive/maximum entropy spectral estimation algorithms, e.g., [1, 10].

II.3. *z-Transforms and Deconvolution*

As in Section II.2, we are given a function $C: \mathbb{R} \to \mathbb{C}$ for which $C(0) = 1$, C vanishes on $(-\infty, 0)$, $C(0+) = 1$, and C is continuous on $(0, \infty)$. Fix any $c > 0$ and define the sequences $\{\alpha_j\}$, $\{\beta_j\}$, $j \geqslant 0$, as in Section II.2. Further, we define the measures

$$\mu_{c,n} = c \sum_{j=0}^{n} \beta_j \delta_{jc}, \qquad n \geqslant 1, \tag{II.3.1}$$

and set $\mu_{c,\infty} = \mu_c$. The *z-transform* of $\mu_{c,n}$ is the analytic function

$$c \sum_{j=0}^{n} \beta_j z^{-j}, \qquad z = re^{ic\theta};$$

this definition is only a slight modification of the usual definition of the z-transform of a sequence, e.g., [32, Chap. 2].

We mention the z-transform in the context of (II.3.1) because of the following fact:

PROPOSITION II.3.1. (a) *There is a measure*

$$\mu_c^{-1} = \frac{1}{c} \sum_{j=0}^{\infty} \gamma_j \delta_{jc}$$

with the properties that $\gamma_0 = 1$, $\mu_c * \mu_c^{-1} = \delta$, *and, for each* $n \geqslant 1$,

$$\gamma_n = \alpha_n - \alpha_{n-1} \qquad \textit{and} \qquad \alpha_n = \gamma_0 + \cdots + \gamma_n. \tag{II.3.2}$$

(b) $\displaystyle\sum_{j=0}^{\infty} \gamma_j z^j = (1-z) \sum_{j=0}^{\infty} \alpha_j z^j.$

Proof. Part (b) is clear from part (a). Part (a) can be computed as in Proposition II.2.1, or by multiplying the power series $\sum \beta_j z^j$ and $\sum \gamma_j z^j$ and setting the product equal to 1. In either case, we obtain the sequence of equations

$$\sum_{j+k=n} \beta_j \gamma_k = 0, \qquad n \geqslant 0,$$

and in this way we compute (II.3.2). Q.E.D.

Analogous to Proposition II.2.1(b), we also have

PROPOSITION II.3.2. *For each* $n \geqslant 1$,

$$\mu_{c,n} * \left(\frac{1}{c} \sum_{j=0}^{n} \gamma_j \delta_{jc} \right) = \delta + \sum_{n \leqslant m \leqslant 2n} \sum_{\substack{j+k=m \\ 0 \leqslant j,k \leqslant n}} \beta_j \gamma_k \delta_{(j+k)c}.$$

EXAMPLE II.3.1. The measures $\mu_{c,n}$ approximate C *weakly* on an interval $(0, \Omega]$, where $n = n(c)$ is chosen so that $n(c)$ is close to Ω, in the following sense: Let F be a continuous function in $[0, \Omega]$. Then

$$|\langle C-\mu_{c,n}, F\rangle| = \left|\int_0^{\Omega} CF - c\sum_{j=0}^{n} \beta_j F(jc)\right|$$
$$= \left|\sum_{j=0}^{n}\left(\int_{jc}^{(j+1)c} CF - c\beta_j F(jc)\right)\right|$$

is small for small c.

Remark II.3.1. The inverse of the measure μ_c in Proposition II.3.1 is a measure; and the inverse of the step function C_c in Proposition II.2.1 is a first-order distribution. Both of these deconvolution procedures are essentially equivalent. In this section, we approximate C weakly but only have to deal with measures for the inverse. In Section II.2, we have approximated C strongly (uniformly) but have to go beyond measures for the inverse. A tradeoff arises in the digital simulation of analog data: the uniform convergence is more reliable than the weak convergence, but the computation of derivatives is less precise than point evaluation.

II.4. *Spectral Estimation and Error Term*

Given the format specified by (II.1.1)–(II.1.4), we shall frequently invoke the following realistic assumption:

ASSUMPTION. The support $\operatorname{supp} S \subseteq [-\Omega, \Omega]$.

We define $C = (1/B^2(0))\, B^2 H$ so that (II.1.2) and the notation from Section II.2 yield

$$S * C_{c,n} = (1/B^2(0))\, E_b + S * (C_{c,n} - (B^2/B^2(0))) \qquad \text{(II.4.1)}$$

for any $c > 0$. In conjunction with the assumption, we set

$$n = n(\Omega, c) = [2\Omega/c],$$

where $[2\Omega/c]$ is the integer part of $2\Omega/c$. For this n we apply Proposition II.2.1(b) to Eq. (II.4.1) and obtain

$$S = (1/B^2(0))\, E_b * C_{c,n}^{-1} + S * (C_{c,n} - (B^2/B^2(0))) * C_{c,n}^{-1} + \mu, \text{ (II.4.2)}$$

where, in light of the assumption, we have

$$\operatorname{supp} \mu \subseteq [\Omega, \infty). \qquad \text{(II.4.3)}$$

If we define

$$K_{b,c} = (1/B^2(0))\, E_b * C_{c,n}^{-1}, \qquad \varepsilon_+ = S * (C_{c,n} - C) * C_{c,n}^{-1},$$

and

$$\varepsilon_- = S * \left(C - \frac{B^2}{B^2(0)}\right) * C_{c,n}^{-1} = -(1/B^2(0)) S * (B^2 H_-) * C_{c,n}^{-1},$$

then (II.4.2) and (II.4.3) combine to give

$$S = K_{b,c} + \varepsilon_+ + \varepsilon_- \qquad \text{on} \quad (-\Omega, \Omega). \tag{II.4.4}$$

Here ε_+ and ε_- are measures and $K_{b,c}$ is a C^∞-function since B^2 is so smooth. Statement (II.4.4) means that $\langle S - (K_{b,c} + \varepsilon_+ + \varepsilon_-), \phi \rangle = 0$ for every continuous function ϕ supported in $(-\Omega, \Omega)$.

We say that $K_{b,c}$ is the *spectral estimator* of S on Ω depending on T, the window b, and $c > 0$.

The function $K_{b,c}$ is computable because of (II.1.2), (II.2.1), (II.2.2), and (II.2.4). Our main concern now is to find T, b, and c so that the *error*, $\varepsilon_+ + \varepsilon_-$, does not detract from the purpose of the spectral estimator $K_{b,c}$ to detect the discrete part of S.

II.5. *A Bound in the Error Term* $\varepsilon_+ + \varepsilon_-$

In this section, we discuss various upper bounds on the measure $\varepsilon_+ + \varepsilon_-$ defined in (II.4.2).

If S is an absolutely continuous function, then $S' \in L^1(\ )$. Also, since $[-\Omega, \Omega]$ is compact, $L^2[-\Omega, \Omega] \subseteq L^1[-\Omega, \Omega]$. We mention this because of the following result:

PROPOSITION II.5.1. *Given the format specified by* (II.1.1)–(II.1.4), *suppose the assumption is valid, and fix* $c > 0$. *If* $S' = f \in L^2(\mathbb{R})$, *then* $\varepsilon_+ + \varepsilon_-$ *is a function and* $\forall \omega \in (-\Omega, \Omega)$,

$$|(\varepsilon_+ + \varepsilon_-)(\omega)| \leqslant \frac{1}{\sqrt{2\pi}} \|f\|_2 \left(\sum_0^n |\alpha_j|\right) \left\{ \left(\int_{-(4\Omega + c)}^0 |B(\lambda)/B(0)|^4 \, d\lambda\right)^{1/2} + \left(\int_0^{2(\Omega + c)} |(C_{c,n} - C)(\lambda)|^2 \, d\lambda\right)^{1/2} \right\}. \tag{II.5.1}$$

Proof. Because of (II.4.2) and the definitions of ε_+ and ε_-, we have

$$\forall \omega \in (-\Omega, \Omega), \qquad (\varepsilon_+ + \varepsilon_-)(\omega) = \sum_{j=0}^n \alpha_j (\tau_{jc} f) * (C_{c,n} - B^2/B^2(0)),$$

where $c > 0$ and $n = [2\Omega/c]$. Thus, for each $\omega \in (-\Omega, \Omega)$, we write

$$(\varepsilon_+ + \varepsilon_-)(\omega) = \frac{1}{2\pi} \sum_{j=0}^{n} \alpha_j \left(\int_{-\infty}^{0} + \int_{0}^{(n+1)c} + \int_{(n+1)c}^{\infty} \right) \times f(\omega - \lambda - jc)(C_{c,n} - B^2/B^2(0))(\lambda)\, d\lambda. \quad \text{(II.5.2)}$$

Note that

$$2\Omega + c \leqslant (n+1)c < 2(\Omega + c) \quad \text{(II.5.3)}$$

by the definition of n.

If $\lambda > (n+1)c$, then $-\lambda - jc < -(n+1)c$ and hence, $\omega - \lambda - jc < \Omega - (n+1)c \leqslant -\Omega - c$ for all $\omega \in (-\Omega, \Omega)$, where the last inequality follows from (II.5.3). Therefore, the integral $\int_{(n+1)c}^{\infty}$ of (II.5.2) is identically zero.

Next, if $\lambda < -(4\Omega + c)$, then $\omega - \lambda - cj > 3\Omega - (j-1)c > 3\Omega - (n-1)c > 3\Omega + c - (2\Omega + c) = \Omega$ for all $\omega \in (-\Omega, \Omega)$, where we have again used (2.5.3). Therefore, the integral $\int_{-\infty}^{0}$ of (II.5.2) can be written as $\int_{-(4\Omega+c)}^{0}$ since $\operatorname{supp} f \subseteq [-\Omega, \Omega]$.

Consequently, if $\omega \in (-\Omega, \Omega)$, then

$$(\varepsilon_+ + \varepsilon_-)(\omega) = \frac{1}{2\pi} \sum_{0}^{n} \alpha_j \left(\int_{-(4\Omega+c)}^{0} + \int_{0}^{(n+1)c} \right) \times f(\omega - \lambda - jc)(C_{c,n} - B^2/B^2(0))(\lambda)\, d\lambda. \quad \text{(II.5.4)}$$

We obtain (II.5.1) by applying Schwarz's inequality to (II.5.4). Q.E.D.

Remark II.5.1. (a) In Proposition II.5.1, we made the hypothesis $S' \in L^2(\hat{\mathbb{R}})$ instead of the more natural hypothesis $S' \in L^1(\hat{\mathbb{R}})$ for the following reason: Our step function deconvolution technique requires that B^2 essentially vanish on a left half line and then *jump* to a nonzero value. Although there is uniform convergence to B^2 by step functions on $[0, \infty)$, for which case the above L^1 condition is satisfactory, we do not have such behavior on $(-\infty, 0]$ since, ultimately, $\operatorname{supp} b$ is compact. On the other hand, the $L^p(-\infty, 0]$ norm of B^2 can be made small for the functions B we consider even though $\sup\{|B(\lambda)| : \lambda \leqslant 0\} > 0$. These functions B are motivated by Example II.1.1, and explicit examples are defined in Section II.7.

(b) For the general situation with no restrictions on S, we have

$$\varepsilon_+ + \varepsilon_- = S * \left(\sum_{0}^{n} \alpha_j \delta_{jc} \right) * \left(\delta + \sum_{0}^{n} (C((k+1)c) - C(kc))\, \delta_{kc} - (2BB')/B^2(0) - C((n+1)c)\, \delta_{(n+1)c} \right).$$

Thus, the total variation of $\varepsilon_+ + \varepsilon_-$ on $(-\Omega, \Omega)$ is bounded by

$$\|S\|_1 \left(\sum_0^n |\alpha_j|\right) \left\|\delta + \sum_0^n (C((k+1)c) - C(kc))\, \delta_{(k+1)c} - (2BB')/B^2(0)\right\|_1, \tag{II.5.5}$$

and this is the analog of the right-hand side of (II.5.1). Consequently, the contribution of the error $\varepsilon_+ + \varepsilon_-$ from the interval $(-\infty, 0)$ is bounded by

$$\|S\|_1 \left(\sum_0^n |\alpha_j|\right) (2/|B^2(0)|) \int_{-\infty}^0 |B(\omega)\, B'(\omega)|\, d\omega.$$

For a fixed $c > 0$, this term can *not* be made small by any window b satisfying (II.1.3) and (II.1.4).

EXAMPLE II.5.1. Let $B^2 = B^2(0)(\chi_{[0,\Gamma)} + \Delta)$, where $\Delta(\omega) = (T\omega + 2\pi)\, \chi_{[-2\pi/T,0)}(\omega)$, and suppose the convolution,

$$S * (\chi_{[0,\Gamma)} + \Delta) = (1/B^2(0))\, E_b,$$

is given data. If $\hat{b} = B$, then b is a better model to approximate (II.1.3) than the function considered in Example II.1.1. The slope T in the definition of B^2 corresponds to the *support* $[-T, T]$ of b, cf. the beginning of Section II.7. Using the calculation and support hypotheses of Example II.1.1, we see that $S = (1/B^2(0))\, E_b' + \varepsilon_+ + \varepsilon_-$ on $(-\Omega, \Omega)$, where $\varepsilon_+ + \varepsilon_- = S' * \Delta$ on $(-\Omega, \Omega)$ for $\Gamma > 2\Omega$ large enough. Thus, if $S' = f \in L^2(\mathbb{R})$, then we have

$$\forall \omega \in (-\Omega, \Omega), \qquad |(\varepsilon_+ + \varepsilon_-)(\omega)| \leqslant \|f\|_2/\sqrt{T}.$$

Even if $f \in L^\infty(\mathbb{R})$, the norm $\|f\|_\infty$ might be unacceptably large for the bound on $\varepsilon_+ + \varepsilon_-$ to be useful, especially if S has peaks.

II.6. *The Hilbert Transform and the Error Term ε_-*

Given $c > 0$, take $\varepsilon > 0$. We showed in Proposition II.5.1 that the contribution of the error $\varepsilon_+ + \varepsilon_-$ from the interval $(-\infty, 0)$ can be made less than ε for an appropriate choice of window, e.g., the windows of Section II.7. In this section, we make some related remarks concerning ε_- and the Hilbert transform. We begin with a computation of ε_-.

PROPOSITION II.6.1. *For a fixed $c > 0$ and a given data window b, we define $B_j^2(\omega) = (B^2 H_-)(\omega - jc)$, and let $\{B_j^2\}'$ be the pointwise derivative of B_j^2 on $\mathbb{R} \setminus \{jc\}$. Then we have*

$$\begin{aligned}\varepsilon_- &= \sum_{j=0}^{n} \alpha_j \tau_{jc} S - \frac{1}{B^2(0)} \sum_{j=0}^{n} \alpha_j S * \{B_j^2\}'(\omega) \\ &= \sum_{j=0}^{n} \alpha_j \tau_{jc} S - \frac{2}{B^2(0)} \sum_{j=0}^{n} \alpha_j \int_{\omega-jc}^{\Omega} (B'B)(\omega-\lambda-jc)\, dS(\lambda).\end{aligned} \tag{II.6.1}$$

Proof. A standard distributional calculation for derivatives of differentiable functions with a jump gives

$$(BH_-)^2 * \delta'_{jc} = \{B_j^2\}' - B^2(0)\, \delta_{jc}.$$

We obtain the result by substituting this calculation into the expression $\varepsilon_- = -(1/B^2(0))(S * (B^2H_-) * C_{c,n}^{-1})$. Q.E.D.

Suppose that, instead of the context of Section II.4, we choose a *data window* b_1 such that

$$(1/2\pi)\, E\{S_{b_1}(\omega, \cdot)\} = S * (B^2H)(\omega), \tag{II.6.2}$$

where $\hat{b}_1 = BH$. The good news is that we can deconvolve the right-hand side of (II.6.2) by the methods we have developed in the previous sections. The bad news is that if we are given a noisy signal $x(t, \alpha)$ on $[-T, T] \times P$ and wish to estimate S, then we can no longer assert that $E\{S_{b_1}(\omega, \cdot)\}$ is a known quantity as we do in (II.1.2). In particular, the support of b_1 is not contained in $[-T, T]$; the extent to which the duration of b_1 lies outside of $[-T, T]$ provides an error term analogous to ε_-.

The following example provides some details, especially about the Hilbert transform, concerning the error term associated with (II.6.2):

EXAMPLE 2.6.1. (a) The Hilbert transform of $b \in L^2(\mathbb{R})$ is $b * \mathrm{pv}(t)$, where pv designates the principal value distribution $\mathrm{pv}(1/u)$. Also, $\hat{H} = \frac{1}{2}\delta + \mathrm{pv}$, e.g., [37, pp. 258–259]. Defining B as $B = \hat{b}$ and setting $\tilde{b} = (BH)\check{}$, we compute

$$\tilde{b}(t) = \int_0^{\infty} B(\omega)\, e^{-i\omega t}\, d\omega = \tfrac{1}{2} b(t) + b * \mathrm{pv}(t), \tag{II.6.3}$$

where, without loss of generality for this calculation, we have assumed that b is even.

(b) Suppose that $\operatorname{supp} b \subseteq [-T, T]$ and $x(t, \alpha)$ is given on $[-T, T] \times P$. Then, recalling that

$$(1/2\pi)\, E\{S_b(\omega, \cdot)\} = (1/2\pi) \int_P \left| \int x(t, \alpha)\, b(t)\, e^{-it\omega}\, dt \right|^2 dp(\alpha) = S * B^2(\omega),$$

we compute

$$S * (B^2H)(\omega) = \int_P \left(\int_{-T}^{T} x(t,\alpha)\,\tilde{b}(t)\,e^{-it\omega}\,dt\right)$$
$$\times \left(\int_{-T}^{T} x(u,\alpha)\,b(u)\,e^{-iu\omega}\,du\right) dp(\alpha)$$
$$+ \int_P \left(\int_{|t|\geqslant T} x(t,\alpha)\,\tilde{b}(t)\,e^{-it\omega}\,dt\right)$$
$$\times \left(\int_{-T}^{T} x(u,\alpha)\,b(u)\,e^{-iu\omega}\,du\right) dp(\alpha), \qquad \text{(II.6.4)}$$

undaunted about switching limits. The quantity,

$$\int_{|t|\geqslant T} x(t,\alpha)\,b(t)\,e^{-it\omega}\,dt,$$

is the only unknown term on the right hand side of (II.6.4), and it can be estimated only when additional information about $x(t,\alpha)$ is provided for $t \notin [-T, T]$.

(c) We now suppose that not only do we know $x(t,\alpha)$ for $t \in [-T, T]$, but that $\sup\{|x(t,\alpha)|: t \in \mathbb{R}, \alpha \in P\} \leqslant K$. With this hypothesis, and since a deconvolution of (II.6.4) involves convolution by δ'_{jc}, we can really only estimate S from (II.6.4) by evaluating

$$\int_T^{\infty} |t\tilde{b}(t)|\,dt.$$

Unfortunately, for most data windows b this quantity is infinite. For example, if $b(t) = \chi_T(t)(\sin t)/t$, then

$$\int_T^{\infty} |t\tilde{b}(t)|\,dt = \int_T^{\infty} \left| t \int_{-T}^{T} \frac{\sin u}{u(t-u)}\,du \right| dt$$
$$= \int_T^{\infty} \left| \pi_T - \int_{-T}^{T} \frac{\sin u}{t+u}\,du \right| dt,$$

where $\pi_T \neq 0$. Now, for $t \notin [-T, T]$, $\int_{-T}^{T} ((\sin u)/(t+u))\,du$ is about

$$2 \sum_{t-T}^{t+T} \log\left(1 - \frac{1}{m}\right)$$

which, in turn, is about $2\log(1 - (2T/(t+T)))$. Thus, $\lim_{t\to\infty} \int_{-T}^{T} ((\sin u)/(t+u))\,du = 0$ and so $\int_T^{\infty} |t\tilde{b}(t)|\,dt = \infty$.

EXAMPLE II.6.2. In light of the problems arising from attempts to adapt (II.6.3) for deconvolution methods, it is natural to pose the following problem: Find $k(t, \omega)$ such that if $S_{b,k}$ is defined by

$$S_{b,k}(\omega, \alpha) = \left| \int_{-T}^{T} b(t)\, x(t, \alpha)\, k(t, \omega)\, dt \right|^2,$$

then

$$(1/2\pi)\, E\{S_{b,k}(\omega, \cdot)\} = S * (HF^2)(\omega),$$

for some known continuous function F for which $F(0) \neq 0$.

We distinguish the special case $k(t, \omega) = e^{i\omega t} K(t)$ from the general case. In either case, the question is to find such an F and k for which

$$\int_{-T}^{T} b(t)\, k(t, \omega)\, e^{-i\gamma t}\, dt = (HF)(\omega - \gamma) \tag{II.6.5}$$

for all $\omega,\ \gamma \in \mathbb{R}$. If the special case were true, then, letting $\lambda = \omega - \gamma$, we deduce from (II.6.5) that $F(0) = 0$, a contradiction. In the general case, if we again consider $\lambda = \omega - \gamma$ in (II.6.5), then $b(t)\, k(t, \omega)\, e^{-i\omega t} = \hat{H} * \hat{F}(t)$, and so we obtain $(d/d\omega)\, k(t, \omega) = itk(t, \omega)$ on every open time interval where b is nonzero. Thus, for *smooth* k, we are reduced to the original Fourier transform situation.

II.7. *Dirichlet Data Windows*

We shall define data windows $b_{T\Gamma}$ satisfying (II.1.3) and (II.1.4) and having properties similar to those in Example II.1.1. These data windows will be used in our algorithm.

We begin with the *Dirichlet kernel*,

$$d_\Gamma(t) = \frac{1}{2\pi} \int_{-\Gamma/2}^{\Gamma/2} e^{it\omega}\, d\omega = \frac{1}{\pi} \frac{\sin(t\Gamma/2)}{t}, \qquad \Gamma > 0,$$

and note that if $T > 0$, then

$$(\chi_T d_\Gamma)\hat{\ }(\omega) = 2\pi d_{2T} * \chi_{\Gamma/2}(\omega). \tag{II.7.1}$$

Equation (II.7.1) reflects a feature of the uncertainty principle: support in the time interval $[-T, T]$ corresponds to *increase* of the Fourier transform in a frequency interval of length $1/T$. In this case, d_{2T} is concentrated in $[-\pi/T, \pi/T] \subseteq \mathbb{R}$ and the sudden increase and decrease occur at $\pm\Gamma/2$.

Setting

$$b_\Gamma(t) = e^{it\Gamma/2} d_\Gamma(t) = \frac{1}{2\pi} \int_0^\Gamma e^{it\omega}\, d\omega = \frac{1}{\pi} e^{it\Gamma/2} \frac{\sin(t\Gamma/2)}{t}$$

and $B_\Gamma = \hat{b}_\Gamma$, we have $B_\Gamma(\omega) = \chi_{[0,\Gamma]}(\omega)$, except at $\omega = 0, \Gamma$.

Next we define the *Dirichlet data window*

$$b_{T\Gamma} = b(T, \Gamma)\, \chi_T b_\Gamma \qquad \text{and} \qquad B_{T\Gamma} = \hat{b}_{T\Gamma}, \tag{II.7.2}$$

where $b(T, \Gamma)$ is chosen so that $(1/(2\pi)) \int B_{T\Gamma}^2 = 1$. It is easy to see that

$$B_{T\Gamma}(0) = \frac{b(T, \Gamma)}{\pi} \int_0^{T\Gamma} \frac{\sin \gamma}{\gamma}\, d\gamma$$

and hence (II.1.3) and (II.1.4) are satisfied. Further, $\overline{b_{T\Gamma}(t)} = b_{T\Gamma}(-t)$ and so $B_{T\Gamma}$ is real valued.

$B_{T\Gamma}$ looks a lot like a continuous version of $(2\pi/\Gamma)^{1/2} \chi_{[0,\Gamma]}$, where the jump at 0 takes place on the interval $[-\pi/T, \pi/T]$, with a similar interval at Γ. The actual details take the following form (keeping in mind that $\int(\sin t/t)\, dt = \int(\sin^2 t/t^2)\, dt = \pi$):

PROPOSITION II.7.1. (a) *For each $T, \Gamma > 0$, we have*

$$b(T, \Gamma) = \pi \sqrt{2} \Big/ \left[\Gamma \int_{-T\Gamma/2}^{T\Gamma/2} \frac{\sin^2 t}{t^2}\, dt \right]^{1/2} \tag{II.7.3}$$

and $\forall \omega \in \hat{\mathbb{R}}$,

$$\begin{aligned} B_{T\Gamma}(\omega)/b(T, \Gamma) &= \frac{1}{\pi} \int_{-T}^{T} \frac{\sin(t\Gamma/2)}{t} e^{it((\Gamma/2) - \omega)}\, dt \\ &= \frac{1}{\pi} \int_{-\omega}^{\Gamma - \omega} \frac{\sin T\gamma}{\gamma}\, d\gamma. \end{aligned} \tag{II.7.4}$$

In particular, we have

$$\forall \omega \in \hat{\mathbb{R}}, \qquad B_{T\Gamma}(\omega)/B_{T\Gamma}(0) = \left(\int_{-T\omega}^{T(\Gamma - \omega)} \frac{\sin \gamma}{\gamma}\, d\gamma \right) \Big/ \left(\int_0^{T\Gamma} \frac{\sin \gamma}{\gamma}\, d\gamma \right). \tag{II.7.5}$$

(b) *Define the constant* $K_\pi = (\int_{-\pi}^{\pi} (\sin \gamma/\gamma)\, d\gamma)/(\int_{-\pi}^{\pi} (\sin^2 t/t^2)\, dt)^{1/2}$ *and the function* $C_{T\Gamma} = (HB_{T\Gamma}^2)/B_{T\Gamma}^2(0)$. *If $T, \Gamma > 0$ and $T\Gamma \geqslant 2\pi$, then for each $\omega \geqslant 0$ we have*

$$|B_{T\Gamma}(\omega)| \leqslant K_\pi \sqrt{2/\Gamma} \tag{II.7.6}$$

and

$$|C_{T\Gamma}(\omega)| \leqslant \left(K_\pi \Big/ \int_0^{2\pi} \frac{\sin\gamma}{\gamma}\,d\gamma\right)^2 \int_{-T\Gamma/2}^{T\Gamma/2} \frac{\sin^2 u}{u^2}\,du$$
$$< \pi \left(K_\pi \Big/ \int_0^{2\pi} \frac{\sin\gamma}{\gamma}\,d\gamma\right)^2. \tag{II.7.7}$$

Proof. (a) By the Plancherel theorem and the definition of $b(T,\Gamma)$, we have

$$1 = (b(T,\Gamma)/\pi)^2 \int_{-T}^{T} \frac{\sin^2(u\Gamma/2)}{u^2}\,du$$

and this leads immediately to (II.7.3) by the substitution $t = u\Gamma/2$.

The first part of (II.7.4) follows from the definition of $b_{T\Gamma}$ in terms of $(\sin(t\Gamma/2))/t$. The second follows since $B_{T\Gamma}(\omega)/b(T,\Gamma)$ is

$$\frac{1}{2\pi}\int_0^\Gamma \int_{-T}^{T} e^{it(\lambda-\omega)}\,dt\,d\lambda = \frac{1}{\pi}\int_0^\Gamma \frac{\sin T(\lambda-\omega)}{\lambda-\omega}\,d\lambda.$$

(b) By (II.7.4) we obtain $|B_{T\Gamma}(\omega)| \leqslant (b(T,\Gamma)/\pi)\int_{-\pi}^{\pi}(\sin\gamma/\gamma)\,d\gamma$.

Combining this inequality with (II.7.3), and using the hypothesis $T\Gamma \geqslant 2\pi$ and the positivity of $(\sin^2 t)/t^2$ we have (II.7.6).

Equation (II.7.6) and the definition of $C_{T\Gamma}$ imply that $|C_{T\Gamma}(\omega)| \leqslant 2K_\pi^2/(\Gamma B_{T\Gamma}^2(0))$. The definition of $B_{T\Gamma}(0)$ and (II.7.3) yield

$$1/B_{T\Gamma}^2(0) = (\Gamma/2)\left\{\int_{-T\Gamma/2}^{T\Gamma/2} \frac{\sin^2 t}{t^2}\,dt \Big/ \left(\int_0^{T\Gamma} \frac{\sin\gamma}{\gamma}\,d\gamma\right)^2\right\}.$$

Thus, for all $\omega \geqslant 0$, we have (II.7.4) since $\int_0^{2\pi}(\sin\gamma/\gamma)\,d\gamma \leqslant \int_0^{T\Gamma}(\sin\gamma/\gamma)\,d\gamma$ when $T\Gamma \geqslant 2\pi$. Q.E.D.

Remark II.7.1. (a) There is a Gibbs phenomenon associated with $B_{T\Gamma}$. This is a necessary byproduct of any data window b we use in which B^2 can be deconvolved in such a way as to reveal the peaks in S.

(b) For large values of $T\Gamma$, we have the following approximations:

$$b(T,\Gamma) \approx \sqrt{2\pi/\Gamma}, \qquad b_{T\Gamma}(0) \approx \sqrt{\Gamma/(2\pi)}, \qquad \text{and} \qquad B_{T\Gamma}(0) \approx \tfrac{1}{2}\sqrt{2\pi/\Gamma}.$$

Because of (II.7.5), we see that $\lim_{T\to\infty} B_{T\Gamma}(\omega) = 0$ for all $\omega < 0$ and

$$\forall \omega \in (0,\Gamma), \qquad \lim_{T\to\infty} C(\omega) = 4. \tag{II.7.8}$$

Also, if $\omega > 0$ and $T > 0$ is fixed, then

$$\lim_{\Gamma\to\infty} C(\omega) = \left(1 + \frac{2}{\pi}\int_{-T\omega}^{0} \frac{\sin\gamma}{\gamma}\,d\gamma\right)^2$$

PROPOSITION II.7.2. *Given the format specified by* (II.2.1)–(II.2.4) *for the function* $C = (HB_{T\Gamma}^2)/B_{T\Gamma}^2(0)$, *fix positive constants* c, ε, *and* Ω, *and let* $C_{c,n}^{-1} = \sum_0^n \alpha_j \delta_{jc}$ *for* $n = [2\Omega/c]$. *Then there is* $T(c, \varepsilon, \Omega)$ *such that for all* $T > T(c, \varepsilon, \Omega)$ *and for all* $\Gamma > 2\Omega$ *we have*

$$\alpha_j = (-3)^j + e_j, \tag{II.7.9}$$

where $|e_j| < \varepsilon$.

Proof. The definition of α_j as given in (II.2.2) is the same as the following formulation: $\alpha_0 = 1$ and $\forall j \geqslant 1$,

$$\alpha_j = 1 + \sum_{k=1}^{j} (-1)^k \sum_{0 \leqslant m_1 < \cdots < m_k < j} C((m_2 - m_1)c) \cdots$$
$$C((m_k - m_{k-1})c)\, C((j - m_k)c), \tag{II.7.10}$$

e.g., [6, Sects. 2 and 6]. Each m_i is an integer and so we are considering terms $C(mc)$, where $m \in [1, j]$ is an integer. Because of our interest in $C_{c,n}$, we take $j \leqslant n$; thus we shall really be looking at the collection $\{C(mc): c \leqslant mc \leqslant 2\Omega\}$.

For $\omega \geqslant 0$, $C(\omega)$ has the form $[\int_{-T\omega}^{T(\Gamma-\omega)} \cdots / \int_0^{T\Gamma} \cdots]^2$, e.g., (II.7.5). Let us write this in the form

$$C(\omega) = \left[\left(1 \Big/ \int_0^{T\Gamma} \frac{\sin\gamma}{\gamma}\,d\gamma\right)\left(\int_0^{T\Gamma} \frac{\sin\gamma}{\gamma}\,d\gamma\right.\right.$$
$$\left.\left. + \int_{-T\omega}^{0} \frac{\sin\gamma}{\gamma}\,d\gamma + \int_{T(\Gamma-\omega)}^{T\Gamma} \frac{\sin\gamma}{\gamma}\,d\gamma\right)\right]^2. \tag{II.7.11}$$

We shall choose the parameters T and Γ in such a way as to ensure that C is practically constant on $[c, 2\Omega]$. We begin by taking

$$\Gamma > 2\Omega,$$

cf. (II.7.8). Thus, for $\omega = mc \in [c, 2\Omega]$, the integral $\int_{T(\Gamma-\omega)}^{T\Gamma}$ of (II.7.11) can be arbitrarily small, uniformly over $[c, 2\Omega]$, for large T. Similarly, the integral $\int_{-T\omega}^{0}$ of (II.7.11) can be made arbitrarily close to $\pi/2$, uniformly over $\omega = mc \in [c, 2\Omega]$, for large T.

To get the epsilonics right, choose $\tilde{\varepsilon} > 0$ so that

$$\forall j = 1, \ldots, n, \qquad \sum_{k=1}^{j} \binom{j}{k} \{\tilde{\varepsilon}\}_k < \varepsilon, \tag{II.7.12}$$

where

$$\{\tilde{\varepsilon}\}_k = \sum_{i=0}^{k-1} \binom{k}{i} 4^i \tilde{\varepsilon}^{k-i}.$$

Then we quantify the *large* T mentioned above by choosing it to have the property that

$$\forall mc \in [c, 2\Omega], \qquad |C(mc) - 4| < \tilde{\varepsilon}. \tag{II.7.13}$$

That this can be effected follows from (II.7.8) and our remarks concerning (II.7.11).

Because of (II.7.13), $C(mc) = 4 + \varepsilon_m$, where $|\varepsilon_m| < \tilde{\varepsilon}$. Thus the positive product $C((m_2 - m_1)c) \cdots C((j - m_k)c)$ of (II.7.10) equals a product of the form $\prod_{i=1}^{k} (4 + \varepsilon_{m_i})$, where $m_i c \in [c, jc]$ and $|\varepsilon_{m_i}| < \tilde{\varepsilon}$. This product is in turn equal to 4^k plus an error term $e(m_1, \ldots, m_k)$ whose absolute value is bounded by $\{\tilde{\varepsilon}\}_k$. Consequently, we compute

$$\begin{aligned}
\alpha_j &= 1 + \sum_{k=1}^{j} (-1)^k \sum_{0 \leqslant m_1 < \cdots < m_k < j} (4^k + e(m_1, \ldots, m_k)) \\
&= \sum_{k=0}^{j} (-1)^k \binom{j}{k} 4^k + \sum_{k=1}^{j} (-1)^k \sum_{0 \leqslant m_1 < \cdots < j} e(m_1, \ldots, m_k) \\
&= (1 + (-4))^j + e_j,
\end{aligned}$$

where, because of (II.7.12), we have $|e_j| < \varepsilon$. Q.E.D.

We should compare the α_j of Proposition II.7.2 with those α_j computed in the number theoretic example of Section II.9.

II.8. *The Algorithm*

In this section, we shall implement the spectral estimator $K_{T\Gamma c} = K_{b,c}$, where $b = b_{T\Gamma}$. The implementation depends on the type of input information in a given problem. Hence, we begin by listing various input situations, and then formulate our algorithm in terms of $K_{T\Gamma c}$ for a digital problem to which the other input situations can be reduced.

Analog Input Situations

(1) Given a noisy signal $x(t, \alpha)$ on $[-T, T] \times P$, where $x(t, \alpha)$ is continuous on $[-T, T]$ for each $\alpha \in P$, assume x is the restriction to $[-T, T] \times P$ of an SSP on $\mathbb{R} \times P$.

(2) Given a single continuous function $x(t)$ on $[0, T]$.

(a) Assume x is a sample function on $[0, T]$ of an SSP (on $\mathbb{R} \times P$) which is strictly stationary and mean and correlation ergodic. (Then we are back to situation 1.)

(b) Letting $T = KU$, we define the functions $x_k(t) = x(t + kU)$, $t \in [0, U]$ and $k = 0, \ldots, K-1$, and the atomic probability space $P = \{\alpha_0, \ldots, \alpha_{K-1}\}$, where $p(\alpha_k) = 1/K$ for each $k = 0, \ldots, K-1$. We set $x(t, \alpha_k) = x_k(t)$ and assume that $x(t, \alpha)$ is the restriction to $[0, U]$ of an SSP on $\mathbb{R} \times P$. This mathematical model is due to Bartlett (*Nature* 1948), e.g., [13, p. 164; 32, p. 548].

Digital Input Situations

The analog situations 1, 2(a), and 2(b) have corresponding digital formulations dealing with signals $x(n)$, $n = 0, \ldots, N-1$.

Remark II.8.1. In the case of analog input, a digital simulation must be effected by proper sampling in order to make use of digital computer methods including the Fast Fourier Transform (FFT). Classically, such sampling is periodic although this need not be the case, e.g., Section II.9. Naturally, the sampling must be done in such a way as to avoid aliasing problems (when a high frequency assumes the alias of a low frequency because of overlapping). The notion of aliasing is a close relative of the uniqueness concept introduced in Remark II.1.1 and the second theorem of F. and M. Riesz quoted in Remark II.1.2(b), e.g., Remark III.3.1(c).

Having listed various input situations and indicated the importance of digital simulation of analog input, we shall now formulate our algorithm for a basic and customary digital problem which, in fact, corresponds to situation 2(b).

Algorithm

Given data $x(t_n)$, where $n = 1, \ldots, N$, $t_1 = 0$, and $t_j - t_{j-1} = d$, we choose positive constants $T = t_M$, Γ, and c. Values of N and M are taken so that $N = KM$ for a positive integer K. We define the functions

$$\forall k = 1, \ldots, K \quad \text{and} \quad \forall m = 1, \ldots, M, \qquad x_k(t_m) = x(t_{(k-1)M+m}),$$

and compute the Fourier transforms

$$\forall k = 1,\dots, K, \qquad F_k(\omega) = \left| \int_0^T b_{T\Gamma}(t)\, x_k(t)\, e^{-it\omega}\, dt \right|^2$$

by means of the FFT. The derivatives F_k' are computed by finite differences and coefficients α_j are generated recursively by the formulas $\alpha_0 = 1$ and

$$\forall j = 1,\dots, [\Gamma/c], \qquad \alpha_j = 1 - \sum_{m=0}^{j-1} \alpha_m B_{T\Gamma}^2((j-m)c)/B_{T\Gamma}^2(0).$$

The *spectral estimator for the power spectrum of* x is

$$K_{T\Gamma c}(\omega) = \frac{1}{KB_{T\Gamma}^2(0)} \sum_{k=1}^{K} \sum_{j=0}^{[\Gamma/c]} \alpha_j F_k'(\omega - jc). \qquad \text{(II.8.1)}$$

Remark II.8.2. Our algorithm begins with Bartlett's periodogram averaging and Welch's (biased) windowing (IEEE(AU)1970). The difference between their approach and ours is that we use windows which allow for effective deconvolution. The choice of T, Γ, and c depends on the specific problem at hand and we shall presently give general guidelines for this choice vis-à-vis minimization of the error $\varepsilon_+ + \varepsilon_-$. The classical notions and techniques used to minimize error are analyzed in [11; 32, Chap. 11; 34, Chap. 12; and 39].

Because of the formulation in Section II.4, the usefulness of (II.8.1) depends on choosing T, Γ, c so that, in the case of a smooth power spectrum S supported by $[-\Omega, \Omega]$,

$$(\varepsilon_+ + \varepsilon_-)(\lambda) = \sum_{j=0}^{2\Omega/c} \alpha_j \int (C_{c,n} - B_{T\Gamma}^2/B_{T\Gamma}^2(0))(\gamma)\, \tau_{jc} S'(\lambda - \gamma)\, d\gamma \qquad \text{(II.8.2)}$$

can be made small on $(-\Omega, \Omega)$. Besides the support assumption on S, various norms of S are also sometimes known and we assume such information in what follows.

EXAMPLE II.8.1. If $S' \in L^\infty(\hat{\mathbb{R}})$ then, because of Eq. (II.8.1), the simplest bound of the error $(\varepsilon_+ + \varepsilon_-)(\lambda)$ on $(-\Omega, \Omega)$ is

$$\|S'\|_\infty \sum_{j=0}^{2\Omega/c} |\alpha_j| \left(\int_{-\infty}^0 |B_{T\Gamma}^2/B_{T\Gamma}^2(0)| + \int_0^\infty |C_{c,n} - B_{T\Gamma}^2/B_{T\Gamma}^2(0)| \right). \qquad \text{(II.8.3)}$$

We shall evaluate the effectiveness of (III.8.3) for various choices of parameters. Recall that each α_j depends on T, Γ, and c.

(a) Suppose c is fixed. Because of the integral $\int_{-\infty}^{0}$ in (II.8.3) and the fact that $\lim_{T\to\infty} B_{T\Gamma}^2(\omega)/B_{T\Gamma}^2(0) = 0$ for $\omega < 0$, it is necessary that T be large to guarantee small error. On the other hand, the integral $\int_0^\infty$ in (II.8.3) is greater than the integral $\int_0^c$, a quantity which approximates $3c$ as $T \to \infty$, e.g., (II.7.8). Consequently, in the case $\Gamma > 2\Omega$, we see that (II.8.3) provides the unsatisfactory error bound of order $c3^{2\Omega/c}$ because of Proposition II.7.2, even though the error contribution from $(-\infty, 0)$ can be made small. Of course, different choices of Γ alleviate this distress to a certain extent as can be seen from (II.7.10).

(b) Suppose T is fixed and we desire to minimize (II.8.3). In the case $\Gamma > 2\Omega$, the only feasible analytic bound of $(\sum |\alpha_j|) \int_{-\infty}^{0}$ (in (II.8.3)) is $K^{2\Omega/c}/(Tc)$ because of (II.7.7), an argument similar to Proposition II.7.2, and the behavior of $B_{T\Gamma}^2$ on $(-\infty, 0]$. (On this last point, we see that $\int_{-\infty}^{0} B_{T\Gamma}^2/B_{T\Gamma}^2(0)$ is about $(\pi/(2T)) \sum_1 1/k^2$.) In order to make $K^{2\Omega/c}/(Tc)$ small, it is necessary to choose c sufficiently large. A relatively sharp choice is $c = 1/(\log T^\beta)$ where $\beta \log K^{2\Omega} \in (0, 1)$. (In fact, in this case we compute

$$K^{2\Omega/c}/(Tc) = (\log T^\beta)(T^{\beta 2\Omega \log K}/T).)$$

Unfortunately, for this value of c, $(\sum |\alpha_j|) \int_0^\infty$ is an upper bound of $(K^{2\Omega/c}/c) \int_{1/T}^{c} |1 - B_{T\Gamma}^2/B_{T\Gamma}^2(0)|$ which, in turn, is about $3K^{2\Omega/c}$. Once again, smaller values of Γ ameliorate the situation by reducing the size of the α_j.

Remark II.8.3. (a) The main observation of Example II.8.1 is the fact that the value of Γ plays a major role in the size of the α_j, and that if we must resort to (II.8.3) to minimize $\varepsilon_+ + \varepsilon_-$, then we must pay careful attention to the value of Γ which we choose. If $\Gamma < 2\Omega$ and c is given, then the α_j will not be maximal as in Proposition II.7.2. Thus we have the following analytic trade-off: good spectral estimation on $[0, \Gamma]$ if and only if Γ is small relative to Ω (so that the α_j are small). Of course, a choice of Ω much larger than supp S involves more calculations than are necessary.

(b) The phenomenon observed in part (a) shows the value of choosing windows b besides those of the form $b_{T\Gamma}$. We desire the properties that supp $b \subseteq [-T, T]$, *that there are sharp rises in* $S * B^2$ *characterizing the peaks in* S, *and that deconvolution of* B^2 *produces peaks corresponding to the sharp rises*. The existence of such windows is intuitively possible and this point of view is the raison d'être for our algorithm.

II.9. *Deconvolution of Step Functions—Arbitrary Sampling*

If, instead of (II.2.3) and (II.2.5), we consider the step function

$$C_\lambda \doteq \sum_{j=0}^{\infty} C(\lambda_j)\chi_{[\lambda_j, \lambda_{j+1})}, \qquad \lambda_0 = 0 < \lambda_1 < \cdots \quad \text{and} \quad \lim_{j\to\infty} \lambda_j = \infty, \tag{II.9.1}$$

then we can construct a distribution C_λ^{-1} for which $C_\lambda * C_\lambda^{-1} = \delta$. This distribution depends on the semigroup generated by the sequence $\{\lambda_j\}$. The theory associated with such deconvolution and an accompanying approximation theory for exponentially weighted L^1-spaces are the aspects of signal analysis studied in [6].

We shall mention two diverse applications where deconvolution of nonperiodic data is required. Higher dimensional digital signal processing abounds with further examples (e.g., [33, Chap. 6] by A. Baggeroer and the whole area of array processing).

Laser-Doppler Anemometers

The laser-Doppler anemometer (LDA) signal processor provides velocity data $x(t)$ for fluid motion at time t. A feature of LDAs is that the fluid is seeded with particles so that lasers can be used to detect the moving particles; this information is then used to compute the velocity of the fluid. If the seeding concentration is sufficiently high, a continuous analog output can be obtained through a digital-to-analog converter [14, p. 468], cf. Remark II.8.1. In many cases of fluid motion, such as that of gases, it is impossible to implement seeding concentrations which are sufficiently high for the data rate to resolve small scale turbulent fluctuations. In these situations, the Burst-type LDA signal processor is used and it provides data at random times $\{t_j\}$, e.g., [14, Sects. 3.1, 3.2, and 3.5]. Any deconvolution required in the subsequent signal analysis can be accomplished by means of the method referred to in our discussion of (II.9.1), where $\{t_j\}$ replaces $\{\lambda_j\}$.

Since the Burst-type LDA necessitates random sampling spectral estimation, spectral estimators for nonperiodic sampling are required. The project of constructing such estimators, with a view to minimizing bias and variance, has begun, e.g., [20].

Riemann Zeta Function

Define the function C of (II.9.1) as

$$C = \sum_{j=0}^{\infty} \chi_{[\log(2j+1), \log 2(j+1))}. \tag{II.9.2}$$

Then $C * C^{-1} = \delta$, where

$$C^{-1} = \sum_{j=0}^{\infty} \alpha_j \delta_{\gamma_j}, \qquad \gamma_j = \log(j+1),$$

and the coefficients α_j are defined as follows: $\alpha_0 = 1$ and

$$\forall n \geqslant 1, \qquad \alpha_n = 1 - \sum_{0 \leqslant m < n} \beta(m, n)\, \alpha_m, \tag{II.9.3}$$

where

$$\beta(m,n)=1, \qquad \text{if} \quad [(n+1)/(m+1)] \quad \text{is odd,}$$
$$=0, \qquad \text{if} \quad [(n+1)/(m+1)] \quad \text{is even,}$$

cf. [6, Sects. 2 and 6]. Here $\beta(m, n)$ is analogous to β_{n-m} in (II.2.1). The two terms are different because in this section we are dealing with the semigroup $\{\log(j+1): j \geqslant 0\}$, whereas in Section II.2 we are dealing with the semigroup $\{jc: j \geqslant 0\}$.

For convenience, we set $a(n)=\alpha_{n-1}$ for all $n \geqslant 1$. Then we can prove that $a: \mathbb{N} \to \mathbb{Z}$ is a multiplicative function ($a(mn)=a(m)\,a(n)$ if $(m,n)=1$), and also that $a(p)=-1$ and $a(p^2)=0$ for odd primes as well as $a(2^n)=2^{n-1}$ for all $n \geqslant 1$. We can also verify that $a(m) \in \{0, \pm 1\}$, if 4 does *not* divide m and that $a(4k)$ is zero or a power of 2. This information is important for two reasons. First, the fact that the α_j are frequently small is useful in spectral estimation, cf. Example II.8.1 and Remark II.8.3(a). Second, if $S(\gamma)=e^{-\alpha\gamma}C(\gamma)$, $\alpha > 0$, then $\hat{S}(t)=F(\alpha, t)\,\zeta(\alpha+it)$, where F never vanishes and ζ is the Riemann zeta function; and, as such, knowledge of the size of the α_j can be utilized to solve various closure problems that arise in analytic number theory, e.g., [6, Sect. 5].

EXAMPLE II.9.1. Equation (II.9.3) is the matrix equation $(\beta_{ij})(\alpha_j)=1$, where $i,j=0, 1,\ldots$, $\beta_{ij}=\beta(i,j)$, $\alpha_0=\beta_{0,0}=1$, and $\beta_{ij}=0$ for $i<j$, cf. Example II.2.1. The jth column of (β_{ij}) for $j \geqslant 1$ consists of alternating sequences of j 0's and j 1's. Thus, the third column is (0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1,...). The zeroth column is (1, 1, 1, 1,...).

III. Unique Power Spectra for Incomplete Signals

III.1. *Uniqueness Criteria for Power Spectra*

We begin by distinguishing between functions which are known to exist and the subclass of such functions whose values can be calculated arbitrarily closely by some numerical analysis or computer technique such as FFT. We refer to elements f of this subclass as *computable functions* even though analytic or computer difficulties, such as transcendental integrands or time/memory considerations, respectively, may make it impossible actually to calculate $f(t)$. Similarly, we define *computable measures*. It is in this context that we referred to *computable functions* several times in Section II, e.g., Sections II.1 and II.4.

A function $x: [-T, T] \times P \to \mathbb{C}$ is an *incomplete signal* if it is the restriction to $[-T, T]$ of some undetermined SSP y. For each such extension

y of x to $\mathbb{R} \times P$, we let R_y, resp. S_y, be the autocorrelation, resp. power spectrum, corresponding to y.

PROPOSITION III.1.1. *Given an incomplete signal* $x: [-T, T] \times P \to \mathbb{C}$, *there is a well-defined computable function* $R_x: (-2T, 2T) \to \mathbb{C}$ *such that* $R_x = R_y$ *on* $(-2T, 2T)$ *for each extension* y *of* x *to* $\mathbb{R} \times P$.

Proof. We take $b = \chi_T$ so that $b * b > 0$ on $(-2T, 2T)$. The periodogram S_b, defined in terms of x and b with range of integration $(-T, T)$, is computable. Also, because of Proposition I.4.1 we have $E\{S_b\} = S_y * B^2$ for each extension y of x; and thus each $S_y * B^2$ is computable.

Consequently, even though R_y is not necessarily computable on all of $\mathbb{R}$, we can compute $R_y(b * b)$; and so, since $b * b$ is known and nonzero on $(-2T, 2T)$, R_y is computable on $(-2T, 2T)$.

For any two eligible extensions y_1 and y_2 of x, we have $(S_{y_1} - S_{y_2}) * B^2 = 0$ on $\hat{\mathbb{R}}$. Thus, $(R_{y_1} - R_{y_2}) b * b = 0$ on $\mathbb{R}$ and so $R_{y_1} = R_{y_2}$ on $(-2T, 2T)$. This completes the proof. Q.E.D.

An open set L is a *Levinson set* if, for any given $\mu \in M(\hat{\mathbb{R}})$, $(\operatorname{supp} \mu) \cap L = \varnothing$ and $\hat{\mu} = 0$ on an interval imply $\mu = 0$. (Recall from Section 2.4 that $\mu \in M(\hat{\mathbb{R}})$ is 0 on an open set U if $\langle \mu, F \rangle = 0$ for each $F \in C_0(\hat{\mathbb{R}})$ for which $\operatorname{supp} F \subseteq U$.)

PROPOSITION III.1.2. *Given an incomplete signal* $x: [-T, T] \times P \to \mathbb{C}$, *if* L *is a Levinson set and* $S_{y_1} - S_{y_2} = 0$ *on* L *for any two extensions* y_1, y_2 *of* x *then there is a uniquely determined positive measure* S *which is the power spectrum of any extension* y *of* x. S *is not necessarily computable.*

Proof. For any extensions y_1 and y_2, we have $R_{y_1} - R_{y_2} = 0$ on $(-2T, 2T)$ by Proposition 3.1.1. The result follows by the definition of a Levinson set. Q.E.D.

Remark III.1.1. If instead of $(R_{y_1} - R_{y_2}) b * b = 0$ for $b = \chi_T$ in the above results, we considered the situation $Q\mu = 0$ for $\mu \in A'(\mathbb{R})$, $\operatorname{supp} \mu$ compact, and $Q \in M(\hat{\mathbb{R}})\hat{}$, then we can still conclude that $Q = 0$ on $\operatorname{supp} \mu$. This is a direct consequence of Wiener's theorem on the inversion of absolutely convergent Fourier series, e.g., [4, Sects. 1.1 and 2.4].

We shall have more to say about Levinson sets in Section III.3. For now, let us point out their relation to the second theorem of F. and M. Riesz by means of the next result. We stated and referred to the Rieszes' theorem in Remarks II.1.2(b) and II.8.1.

PROPOSITION III.1.3. *Given an incomplete signal* $x: [-T, T] \times P \to \mathbb{C}$, *if the condition,*

$$\exists \Omega > 0 \quad \text{such that} \quad \forall y, \quad \operatorname{supp} S_y \subseteq [-\Omega, \Omega], \tag{III.1.1}$$

is satisfied, where y is an extension of x to $\mathbb{R} \times P$, *then there is a uniquely determined positive measure S which is the power spectrum of every extension y of x. The measure S is not necessarily computable.*

Proof. Let y_1 and y_2 be extensions of x. Then $R_{12} = R_{y_1} - R_{y_2}$ on $(-2T, 2T)$ by Proposition III.1.1 and $S_{12} = S_{y_1} - S_{y_2} = 0$ on $(-\infty, -\Omega)$ by (III.1.1). If $R_{12} \in L^1(\mathbb{R})$, we conclude that $S_{12} = 0$ by the Rieszes' theorem.

If $R_{12} \notin L^1(\mathbb{R})$, we adjust the above argument as follows: Take the Fejér kernel $\{f_{1/n}\}$ defined by the property that $f_{1/n}$ is the isosceles triangle centered at the origin having height $2\pi n$ and base $2/n$, cf. Remark I.3.2. We have $R_{12} f_{1/n} = 0$ on $(-2T, 2T)$ and $S_{12} * \hat{f}_{1/n} = 0$ on a half line. Further, $R_{12} f_{1/n} \in L^1(\mathbb{R})$ since R_{12} is bounded and $f_{1/n} \in L^1(\mathbb{R})$; we cannot use the Dirichlet kernel of Section II.7 since $d_T \notin L^1(\mathbb{R})$. By the Rieszes' theorem, we have $S_{12} * \hat{f}_{1/n}$ identically zero; and since $\lim_{n \to \infty} \hat{f}_{1/n} = \delta$ in the weak $*$ topology $\sigma(M(\hat{\mathbb{R}}), C_0(\hat{\mathbb{R}}))$, we conclude that $S_{12} = 0$. Q.E.D.

III.2. *The Uniqueness Point of View for Spectral Estimation*

III.2.1 *Point of View*

Given an incomplete signal $x\colon [-T, T] \times P \to \mathbb{C}$, if we could record the signal as it runs its course over all time, then, a fortiori, we would see that x is the restriction to $[-T, T] \times P$ of a unique noisy signal on $\mathbb{R} \times P$ with a well-defined, although not necessarily computable, power spectrum S.

Taking this (uniqueness) point of view, we can ask the following question at time T, since we do not have the luxury of a complete reading of the signal over all time: Are there constraints built into the incomplete signal and physical system which allow us to verify the uniqueness of S? Because of our point of view, such constraints are not a priori required to ensure uniqueness. Without them, however, the data at time T could presumably point to many different power spectra; and since we would like to make decisions at time T about periodicities in the signal, it is important to have the right S to estimate. The analytic hypotheses about Levinson sets and power spectrum support in Propositions II.1.2 and II.1.3 are examples of such constraints. Such hypotheses are not unrealistic since physical systems can only deal with band limited autocorrelations.

It is well to note that the algorithm of Section II depends on the point of view we have just described, as opposed to the modeling and nonuniqueness inherent in MEM which we shall outline below. In fact, our algorithm depends on the formula $E\{S_b\} = S * B^2$ derived in Section I.4; this formula asserts the existence of a well-defined power spectrum S based on an incomplete signal as input. Of course, there are theoretical and technical problems about extricating S from the formula. To explain these problems, we note that an objection to the periodogram is that an incomplete signal

$x: [-T, T] \times P \to \mathbb{C}$ is really defined as 0 for $t \notin [-T, T]$, cf. our discussion of MEM below. This deterministic definition is compensated for in our algorithm by complete flexibility in the choice of windows supported by $[-T, T]$. This flexibility allows us to acquire maximum information about S from the incomplete signal by means of the deconvolution of optimal spectral windows B^2. This is as much as one can expect from any algorithm.

EXAMPLE III.2.1. As we mentioned in Remark II.8.3(b), the windows of Section II.7 are not the only choices for our frequency-deconvolution. Using Chebyshev polynomials, Tseng *et al.* [38] have devised a technique for generating a class of data windows b having sidelobe control parameters for the corresponding spectral windows B. A particular example is $B = b_r$, defined in Section II.7.

III.2.2 *Maximum Entropy Method* (MEM)

Given an incomplete signal x on $[-T, T] \times P$, we let y be an extension of x to $\mathbb{R} \times P$ and we let R_y, resp. S_y, be the autocorrelation, resp. power spectrum, of y. If each S_y is a *function* and each R_y takes the same computable values on $[-T, T]$, then the maximization of a certain canonical entropy integral $H(S_y)$, over all extensions y of x, yields the MEM spectral estimator K_{MEM}, e.g., [33, pp. 410–416 by A. Baggeroer; 34, pp. 248–251; 10]. Since entropy is a measure of disorder in a system, K_{MEM} represents maximum uncertainty with regard to what we do not know about the system, whereas it depends on all the known autocorrelation data. Burg (1967) introduced the entropy approach and developed a method for computing the prediction error coefficients in K_{MEM} when K_{MEM} is a digital estimator. Fougere [19] has refined this method by means of an iterative procedure involving nonlinear optimization. This method has had many successes. Some of its weaknesses include extensions to higher dimensions and problems involving nonperiodic sampling.

In applications of MEM, $2M + 1$ periodically sampled autocorrelation data are given on $[-T, T]$ and K_{MEM} is shown to be a rational function in $e^{it\omega/M}$ of the form $1/P$, where $P \geqslant 0$ is a polynomial of degree $2M + 1$. The point of view of MEM is to choose periodic digital autocorrelation data for R outside of $[-T, T]$ in such a way that the resulting power spectrum is a periodic rational function, a restriction motivated by entropy considerations, e.g., [34, p. 250].

Classical periodogram spectral estimation specifies R to be 0 outside of the given data interval $[-T, T]$, but has an open mind about the ultimate shape of S. On the other hand, MEM molds the values of R outside of the given data interval $[-T, T]$ so as to compute S as a rational function. As such, the uniqueness point of view espoused above in Section III.2.1 is essentially different from that of MEM; it also neutralizes the argument against the classical periodogram, which specifies R outside $[-T, T]$, because the associated algorithm has the capability of deconvolving over large families of spectral windows.

III.2.3 *Algorithm Error*

We are given an incomplete signal $x: [-T, T] \times P \to \mathbb{C}$. Note that the uniqueness criteria of Section III.1 do *not* depend on the length of T. Because of Remark II.8.3(b), the length of T does affect the effectiveness of the estimator $K_{T\Gamma_c}$; T must be sufficiently large to resolve two closely situated peaks or to resolve peaks of a given height. In both these cases, T can be quantified in terms of the distance between peaks or of the given height, respectively. Recall that Example II.8.1 discusses algorithm error when input data is supplied on all of $[-T, T] \times P$.

As we pointed out in analog input situation 2(b) of Section II.8, we are frequently only given a single sample function x on $[-T, T]$ instead of information on all of $[-T, T] \times P$. If $T = KU$ in this case, then large K establishes a reliable probability space and cuts down the variance of the associated Bartlett–Welch estimator, e.g., [32, pp. 548–554]. In fact, large K and hence large T reduce both bias and variance. Since our spectral estimator of Section II is unbiased, we present another type of bias result in Section III.4.

Both the case of a single analog sample function or the digital simulation thereof produce errors in estimation prior to dealing with the input domain $[-T, T] \times P$ of our algorithm. These errors are still dealt with by the usual methods.

EXAMPLE III.2.2. If the uniqueness point of view is spotlighted on an incomplete signal $x: [-T, T] \times P \to \mathbb{C}$ whose (unique) power spectrum S contains no δ's, then the noisy extension of x to $\mathbb{R} \times P$ is an *ergodic process* by Maruyama's theorem (1949), e.g., Grenander (*Arkiv Mat.*, 1950) and [18, pp. 76–78]; this is theoretically an uninteresting situation in spectral estimation. On the other hand, the hypothesis $S' \in L^\infty(\hat{\mathbb{R}})$ of Example II.8.1 is not realistic if S has peaks, since $\|S'\|_\infty$ might be too large to control $\varepsilon_+ + \varepsilon_-$, even if it were possible to make effective estimates on the remainder of (II.8.3), cf. Example II.5.1.

III.3. *Levinson Sets and Uniqueness Theorems*

There are two types of uniqueness theorems we shall discuss which are related to errors in digital simulation (Section II.8) and to our uniqueness point of view (Sections III.1 and III.2), respectively. The former were originally stated by Cauchy in the 19th century, proved by Whittaker in 1915, applied by Shannon in the 1940s (Shannon sampling), and elevated to the supernatural by Beurling and Malliavin in 1967, e.g., [36] and the essay by Koosis in Volume 2 of Wiener's *Oeuvres*, pages 892–908. The latter provide the deepest insights into the relation between the domain of a function x and the domain of $\hat{x} = X$; the most recent results are due to Benedicks [7], Beurling [9], and de Branges [12]. These results are related to spectral synthesis and the uncertainty principle, and they are required in Section IV for the solution of Tauberian minimization problems in terms of information from the distant past.

The classic work by Levinson [28] makes fundamental contributions to both types of theorems; and the bibliographies in the above references give a good sampling of the extensive literature on the subject. The first kind of uniqueness result depends on *band limited functions* x, i.e., $\operatorname{supp} X$ is compact for $\hat{x} = X$. This is the subject of Proposition III.3.1 and Remark III.3.1. The second kind of uniqueness result depends on signals x with *time gaps*, i.e., the set $\{t : x(t) = 0\}$ contains an interval. This is the subject of Proposition III.3.2 and the remainder of Section III.3; this material motivated our definition of Levinson set in Section III.1. For band limited functions, we obtain uniqueness given sample values of x. For time gap functions, we obtain uniqueness given support properties of X.

The following result is a standard version of Shannon sampling:

PROPOSITION III.3.1. *Suppose $X \in L^1(\mathbb{R})$ is supported by $[\Omega_0, \Omega_0 + 2\Omega]$.*

(a) *If $T = \pi/\Omega$ and $x = \hat{X}^{-1}$, then*

$$\forall t \in \mathbb{R}, \qquad x(t) = \sum_{n=-\infty}^{\infty} x(nT) \frac{\sin[(t - nT)\Omega]}{(t - nT)\Omega}. \tag{III.3.1}$$

(b) *If $T \leqslant \pi/\Omega$, $x = \hat{X}^{-1}$, and $x(t_0 + nT) = 0$ for each $n \in \mathbb{Z}$, then $x = 0$.*

Proof. This proof is due to Pollard and Shisha (1972). Without loss of generality, suppose $\Omega_0 = -\Omega$ and $t \notin T\mathbb{Z}$. Then we have

$$x(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} X(\omega)\, e^{i\omega t}\, d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \left(\sum a(n)\, e^{-in\omega\pi/\Omega} \right) e^{i\omega t}\, d\omega$$

$$= \sum \frac{a(n)}{2\pi} \int_{-\Omega}^{\Omega} e^{i\omega(t-nT)}\, d\omega; \tag{III.3.2}$$

this is the right-hand side of (III.3.1) since the Fourier coefficient $a(n)$ equals $Tx(nT)$ for each n. Note that the calculation depends on an interplay between Fourier series and transforms. Also, the term-by-term integration follows from the classical fact (Titchmarsh's *Theory of Functions*, 2nd ed., pp. 419–422) that Fourier series of integrable functions $X(\omega)$ on $[-\Omega, \Omega)$

multiplied by functions of bounded variation $e^{i\omega t}$ on $[-\Omega, \Omega)$ can be integrated term by term.

Part (b) is clear from (III.3.1). Q.E.D.

Parts (a) and (b) of Proposition III.3.1 show the essential connection between closure and uniqueness theorems, cf. [4, pp. 100–101].

Remark III.3.1. (a) An extensive survey of sampling theorems since Shannon's work is due to Jerri (*Proc. IEEE*, 1977). There are sampling theorems over zeros of Bessel functions and for locally compact abelian groups (Kluvanek, 1965). There are also sampling theorems for functions X with unbounded support (Boas in *Tohoku*, 1972, where Poisson summation plays a role).

(b) Generally, sampling theorems that sample nonperiodic data are difficult, and it is this problem which is the subject matter of Beurling and Malliavin (1967). For a given sequence $\{t_n\}$ of sampling times, they give necessary and sufficient conditions for the span of $\{\exp it_n\omega\}$ to be dense in $L^2[-\Omega, \Omega]$. The fact that such a result corresponds to a sampling theorem is accounted for by the following observation: Suppose $X \in L^2(\mathbb{R})$, $\operatorname{supp} X \subseteq [-\Omega, \Omega]$, $x = \hat{X}^{-1}$, and $\lim_n \|X - P_n\|_{L^2[-\Omega,\Omega]} = 0$, where $P_n(\gamma) = \sum_{|m| \leqslant n} a(m) \exp(-it_m\gamma)$. If $\hat{p}_n = P_n$, then $p_n = \sum_{|m| \leqslant n} a(m)\, \delta_{t_m}$ and $\lim_n \|x - (\chi_\Omega P_n)\hat{}\,\|_{L^2(\mathbb{R})} = 0$. Note that $(\chi_\Omega P_n)\hat{}$ has the same form as the right-hand side of (III.3.1)! An approximate identity argument also allows us to write $a(n)$ in terms of $x(t_n)$ just as we did in Proposition III.3.1.

(c) We mentioned aliasing problems with regard to digital simulation in Remark II.8.1; this is closely related to Shannon sampling in the following way: Suppose we are given a band limited analog sample function $x(t)$ for which $\operatorname{supp} X \subseteq [-\Omega, \Omega]$, $\hat{x} = X \in L^1(\mathbb{R})$. If the digital sampling period of x is T and $T \leqslant \pi/\Omega$ then we have (III.3.1). If $T > \pi/\Omega$, then the periodic repetitions of X on $\mathbb{R}$, which occur by treating the Fourier series of X in (III.3.2), will overlap. This situation is the phenomenon of *aliasing*, cf. [11]. The *Nyquist rate* is the sampling rate $1/T = \Omega/\pi$.

The following proposition is the primordial time gap result:

PROPOSITION III.3.2. *Suppose $X \in L^1(\mathbb{R})$ satisfies the condition, $|X(\gamma)| \leqslant B \exp b\gamma$, for all $\gamma \geqslant 0$ and a fixed $b < 0$. If $x = \hat{X}^{-1}$ and $x = 0$ on an interval, then $x = 0$.*

Proof. We begin with the estimate

$$|x(-z)| \leqslant \frac{1}{2\pi} \int_{-\infty}^{0} |X(\gamma)|\, e^{u\gamma}\, d\gamma + \frac{B}{2\pi} \int_{0}^{\infty} e^{b\gamma} e^{u\gamma}\, d\gamma$$

$$\leqslant \|X\|_1 + \frac{B}{2\pi(u+b)} < \infty,$$

where $z = t + iu$ and $0 \leqslant u < -b$. Similarly, $x'(-z)$ exists in the strip $0 < u < -b$. Thus, $x(-z)$ is an analytic function in a strip. Consequently, $x(t)$ is the boundary value of an analytic function; hence, $x = 0$ since $x = 0$ on an interval. Q.E.D.

In order to generalize this result, Levinson essentially implemented the quasi-analytic condition,

$$\int_{-\infty}^{\infty} \frac{\log^+ |K(\gamma)|}{1+\gamma^2}\, d\gamma = \infty, \tag{III.3.3}$$

where $\log^+ y = \max\{0, \log y\}$ for $y \geqslant 0$. Condition (III.3.3) plays a role in several other uniqueness-closure problems including the spectral analysis of weighted L^1-spaces and the Bernstein approximation problem.

The former deals with characterizing invariant subspaces, as one does in $L^1(\mathbb{R})$ by means of Wiener's Tauberian theorem, e.g., Section IV and [4, pp. 25–26, 29–30; 6, Introduction]; Beurling and Domar have made major contributions to this problem. The latter was solved by Pollard [35] and because of his role in the uniqueness theory of this section, we describe his solution. Let $\Phi > 0$ be a continuous function such that $\lim_{|\gamma| \to \infty} \gamma^n/\Phi(\gamma) = 0$ for each $n \geqslant 0$; the span of the sequence $\{\gamma^n/\Phi(\gamma): n = 0, 1, \ldots\}$ is $\| \ \|_\infty$-dense in $C_0(\hat{\mathbb{R}})$ if and only if

$$\sup_K \int_{-\infty}^{\infty} \frac{\log^+ |K(\gamma)|}{1+\gamma^2}\, d\gamma = \infty,$$

where K is a real polynomial for which $|K(\gamma)| < \Phi(\gamma)$. He proved this result by introducing an important lemma about entire functions of exponential type. Later, de Branges (1958) used this lemma in his more general reformulation and new proof of Levinson's theorem (1936), e.g., [12, pp. 251ff.], cf. [28, Chapter 5].

THEOREM III.3.1 (Levinson). *Let $K \geqslant 1$ be a continuous function on $\hat{\mathbb{R}}$ for which $\log K$ is uniformly continuous and* (III.3.3) *holds. If $S \in M(\hat{\mathbb{R}})$ satisfies the conditions that $\int K d|S| < \infty$ and $\hat{S} = 0$ on an interval, then $S = 0$.*

EXAMPLE III.3.1. Proposition III.3.2 is a special case of Theorem III.3.1. To see this, take X as in Proposition III.3.2. Choose the function K to be 1 if $\gamma < 0$ and $e^{c\gamma}$ if $\gamma \geqslant 0$, where $0 < c < -b$, and let $S = X$. Then the conditions of Theorem III.3.1 are satisfied and so $X = 0$.

For entire functions K of exponential type, condition (III.3.3) has been characterized by Beurling and Malliavin (1962). Recall that an entire function K is of *exponential type* if there are constants B and b such that

$|K(z)| \leqslant B \exp b\,|z|$ for all $z \in \mathbb{C}$. Beurling and Malliavin proved the difficult necessary conditions (for (III.3.4)) in the following theorem (e.g., [24]):

THEOREM III.3.2. *Let K be an entire function of exponential type. Then the function K satisfies the condition*

$$\int_{\infty}^{\infty} \frac{\log^{+} |K(\gamma)|}{1+\gamma^2}\, d\gamma < \infty, \tag{III.3.4}$$

if and only if for each $b > 0$ there is a nonzero entire function L_b of exponential type $\leqslant b$ for which $(1 + |K|)\,|L_b| \in L^{\infty}(\hat{\mathbb{R}})$, cf. the Paley–Wiener condition in Section IV.1.

Also, in their important 1967 paper which we have already mentioned, Beurling and Malliavin are concerned mainly with the distribution of zeros of the entire functions K of exponential type which satisfy (III.3.4).

EXAMPLE III.3.2. It is possible to construct Levinson sets by a careful analysis of (III.3.3). Explicit examples of Levinson sets have been constructed by Beurling [9] and deBranges [12, p. 280]. An important recent analysis of such sets is due to Benedicks [7], cf. H. S. Shapiro (*Proc. AMS*, 1973) as well as Ehrenpreis' Watergate problem (*Bull. Sci. Math.*, 1981).

Remark III.3.2. There are two natural problems which arise from the discussion in this section.

(a) Besides results such as Theorem III.3.1 and Example III.3.2, the subject of spectral estimation also asks for conclusions similar to uniqueness which depend on the *size* of the time gap and conditions *weaker* than (III.3.3).

(b) Theorem III.3.1 and all known extensions depend on complex analysis proofs or properties of harmonic functions, cf. Doss' real proof of the Rieszes' theorem for the torus (*Proc. AMS*, 1981). Multidimensional spectral estimation problems, such as those arising in 2-dimensional image processing, sometimes warrant multidimensional generalizations of Theorem III.3.1 depending on real-variable proofs not involving harmonic functions. One possible strategy is a systematic use of Wiener's Tauberian theorem and the uncertainty principle.

III.4. *Resolution of Spectral Estimators for Ergodic Processes*

We saw in Section III.1 that if $x\colon [-T, T] \times P \to \mathbb{C}$ is an incomplete signal then power spectrum support hypotheses yield uniqueness. If, in addition, power spectrum smoothness hypotheses are available, then we shall prove in

Theorem III.4.1 that a modified Bartlett–Welch estimator provides effective spectrum resolution even when T is small. Some of this smoothness is reflected in the hypothesis that the extensions of x are ergodic processes so that S is a continuous measure, e.g., Example III.2.2.

We shall formulate the results of this section in terms of convolutions $S * B^2 * V$, thus connecting the material here with the classical calculation showing the consistency of Bartlett–Welch estimation, e.g., [34, pp. 389ff.]. Here $B \in L^2(\mathbb{R})$ will be real, even, and satisfy the condition $(1/(2\pi)) \int B^2(\omega)\, d\omega = 1$, and $\mathscr{V}$ will be the set of compactly supported functions $V \in A(\mathbb{R})$ which are real, even, continuously differentiable, and satisfy the conditions $(1/(2\pi)) \int V = 1$ and $V(0) \geqslant |V(\omega)|$ on $\mathbb{R}$. A function $F\colon \mathbb{R} \to \mathbb{C}$ is *Lipschitz* of order α and is an element of $\mathrm{Lip}_C\, \alpha$, $\alpha > 0$, if

$$\forall \lambda > 0 \quad \text{and} \quad \forall \omega \in \mathbb{R}, \qquad |F(\omega - \lambda) - F(\omega + \lambda)| \leqslant C\lambda^{\alpha}.$$

If $F \in \mathrm{Lip}_C\, 1$, then F is absolutely continuous and so $F' \in L^1_{\mathrm{loc}}(\mathbb{R})$.

PROPOSITION III.4.1. *Given $V \in \mathscr{V}$ and B, and letting $S \in M(\mathbb{R})$ be positive, we have*

$$\|S * B^2 * V - S * V\|_\infty \leqslant 3V(0)\, \|S\|_1, \tag{III.4.1}$$

where $B^2 * V \in C_0(\mathbb{R})$ and $S * V$ is continuous and bounded.

Proof. Let $F' = S$, where $F(-\infty) = 0$, $F(+\infty)/(2\pi) = \|S\|_1$, and F is an increasing function. We use the mean value theorem for integrals to compute

$$\begin{aligned} &|S * B^2 * V(\omega) - S * V(\omega)| \\ &= \frac{1}{2\pi} \left| \int F(\omega - \gamma)(B^2 * V'(\gamma) - V'(\gamma))\, d\gamma \right| \\ &= \varlimsup_{\Omega \to \infty} (1/2\pi)\, |F(\omega - (\Omega -))(B^2 * V(\Omega) - V(\Omega) - B^2 * V(\xi_\Omega) + V(\xi_\Omega)) \\ &\quad + F(\omega - ((-\Omega)+))(B^2 * V(\xi_\Omega) - V(\xi_\Omega) - B^2 * V(-\Omega) + V(-\Omega))|, \end{aligned}$$

where $-\Omega < \xi_\Omega < \Omega$. Equation (III.4.1) is a consequence of this last expression. Q.E.D.

THEOREM III.4.1. *Given B and letting $S \in M(\mathbb{R})$ be positive.*

(a) *If S is continuous then*

$$\forall \varepsilon > 0\ \exists \lambda > 0 \quad \textit{such that} \quad \|S * B^2 * V - S * V\|_\infty \leqslant (\varepsilon V(0)/2\pi)((8\pi + 1)/2\pi) \tag{III.4.2}$$

for all $V \in \mathscr{V}$ for which supp $V \subseteq [-\lambda, \lambda]$.

(b) *If* $F' = S$ *and* $F \in \mathrm{Lip}_C\,\alpha$, $\alpha > 0$, *then* S *is continuous and*

$$\forall \varepsilon > 0 \quad \textit{and} \quad \forall r > 0, \qquad \|S * B^2 * V - S * V\|_\infty < ((8\pi + 1)/2\pi)\, C^{1/\alpha} \varepsilon^{1-1/\alpha} + r \tag{III.4.3}$$

for all $V \in \mathscr{V}$ *for which* $\operatorname{supp} V \subseteq [-(\varepsilon/C)^{1/\alpha}, (\varepsilon/C)^{1/\alpha}]$.

Proof. (a) Since $F' = S$ is continuous, we know that the function F is continuous. Since F is bounded and increasing, a routine calculation shows further that F is uniformly continuous on $\mathbb{R}$. Thus for a given $\varepsilon > 0$ in (III.4.2), we choose $\lambda > 0$ such that $|\omega - \gamma| < 2\lambda$ implies $|F(\omega) - F(\gamma)| < \varepsilon$. Now take any $V \in \mathscr{V}$ for which $\operatorname{supp} V \subseteq [-\lambda, \lambda]$.

For any $\omega \in \mathbb{R}$ we have the inequality

$$|S * B^2 * V(\omega) - S * V(\omega)| \leqslant \frac{1}{2\pi} \left| \int_{|\gamma| \leqslant \lambda} F(\omega - \gamma)(B^2 * V'(\gamma) - V'(\gamma))\, d\gamma \right| + \frac{1}{2\pi} \left| \int_{|\gamma| \geqslant \lambda} F(\omega - \gamma)\, B^2 * V'(\gamma) d\gamma \right|. \tag{III.4.4}$$

To prove the result, we shall estimate both terms on the right hand side of (III.4.4) for a fixed $\omega \in \mathbb{R}$.

The first term is

$$\begin{aligned} &(1/2\pi)\, |F(\omega + \lambda)(B^2 * V(\xi) - V(\xi) - B^2 * V(-\lambda)) \\ &\quad + F(\omega - \lambda)(B^2 * V(\lambda) - B^2 * V(\xi) + V(\xi))| \\ &\leqslant \frac{2V(0)\varepsilon}{2\pi} + \frac{1}{(2\pi)^2} \left| (F(\omega + \lambda) - F(\omega - \lambda)) \int B^2(\lambda - \gamma)\, V(\gamma)\, d\gamma \right|. \end{aligned} \tag{III.4.5}$$

In (III.4.5) we have used the mean value theorem for integrals since F is monotonic, as well as the definition of $\mathscr{V}$ and the fact that B and V are even: $\int B^2(-\lambda - \gamma)\, V(\gamma)\, d\gamma = \int B^2(\lambda - \gamma)\, V(\gamma)\, d\gamma$. Because of (III.4.5), we have

$$\frac{1}{2\pi} \left| \int_{|\gamma| \leqslant \lambda} F(\omega - \gamma)(B^2 * V'(\gamma) - V'(\gamma))\, d\gamma \right| \leqslant \frac{3\varepsilon V(0)}{2\pi}. \tag{III.4.6}$$

The second term on the right-hand side of (III.4.4) is

$$\frac{1}{4\pi^2} \left| \int B^2(\eta) \left[\int_\lambda^\infty F(\omega - \gamma)\, V'(\gamma - \eta)\, d\gamma + \int_{-\infty}^{-\lambda} F(\omega - \gamma)\, V'(\gamma - \eta)\, d\gamma \right] d\eta \right|. \tag{III.4.7}$$

The term in square brackets in (III.4.7) is

$$\int_{\lambda-\eta}^{\lambda} V'(\gamma)\,F(\omega-\eta-\gamma)\,d\gamma + \int_{-\lambda}^{-\lambda-\eta} V'(\gamma)\,F(\omega-\eta-\gamma)\,d\gamma \quad \text{(III.4.8)}$$

by the support property of V. In (III.4.8) we have

$$\forall \eta \leqslant 0, \qquad \int_{\lambda-\eta}^{\lambda} V'(\gamma)\,F(\omega-\eta-\gamma)\,d\gamma = 0$$

and

$$\forall \eta \geqslant 0, \qquad \int_{-\lambda}^{-\lambda-\eta} V'(\gamma)\,F(\omega-\eta-\gamma)\,d\gamma = 0.$$

Thus, using the hypothesis that B is even, a straightforward manipulation shows that the integral in (III.4.7) is

$$\int_0^{2\lambda} + \int_{2\lambda}^{\infty} B^2(\eta) \left\{ \int_{\lambda-\eta}^{\lambda} V'(\gamma)\,F(\omega-\eta-\gamma)\,d\gamma \right.$$
$$\left. + \int_{-\lambda}^{-\lambda+\eta} V'(\gamma)\,F(\omega+\eta-\gamma)\,d\gamma \right\} d\eta. \qquad \text{(III.4.9)}$$

(i) If $\eta \geqslant 2\lambda$, then the term in braces in (III.4.9) is

$$F(\omega-\eta+\lambda)\,V(\xi_1) + F(\omega-\eta-\lambda)(-V(\xi_1)) + F(\omega+\eta+\lambda)\,V(\xi_2)$$
$$+ F(\omega+\eta-\lambda)(-V(\xi_2)), \qquad -\lambda < \xi_1, \quad \xi_2 < \lambda,$$

where once again we have used the monotonicity of F and the mean value theorem for integrals. Substituting this information into (III.4.9), we obtain

$$\left| \int_{2\lambda}^{\infty} B^2(\eta)\{\ \}\,d\eta \right| \leqslant 2\varepsilon V(0) \int_{2\lambda}^{\infty} B^2(\eta)\,d\eta. \qquad \text{(III.4.10)}$$

(ii) If $\eta \in [0, 2\lambda]$ then the term in braces in (III.4.9) is

$$V(\xi_1)(F(\omega-\lambda) - F(\omega-\eta-\lambda)) + V(\lambda-\eta)(F(\omega+\lambda) - F(\omega-\lambda))$$
$$+ V(\xi_2)(F(\omega+\eta+\lambda) - F(\omega+\lambda)),$$
$$\lambda - \eta < \xi_1 < \lambda \quad \text{and} \quad -\lambda < \xi_2 < -\lambda + \eta;$$

and substituting this into (III.4.9) yields

$$\left| \int_0^{2\lambda} B^2(\eta)\{\ \}\,d\eta \right| \leqslant 3\varepsilon V(0) \int_0^{2\lambda} B^2(\eta)\,d\eta. \qquad \text{(III.4.11)}$$

Equations (III.4.10) and (III.4.11) allow us to bound (III.4.7) by $\varepsilon V(0)(1+(1/2\pi))/(2\pi)$; and combining this with (III.4.6), we obtain (III.4.2).

(b) Given ε and r, set $\delta = 2(\varepsilon/C)^{1/\alpha}$. Since $F \in \mathrm{Lip}_C\,\alpha$, we know that S is continuous because F is a continuous function. The right-hand side of (III.4.3) reflects a search for the smallest possible value of $V(0)$ in part (a). By our smoothness hypothesis, we have

$$\forall 0 < 2\lambda \leqslant \delta \quad \text{and} \quad \forall \omega, \qquad |F(\omega-\lambda) - F(\omega+\lambda)| \leqslant C\lambda^\alpha \leqslant \varepsilon.$$

Therefore, the λ chosen in part (a) is $\delta/2 = (\varepsilon/C)^{1/\alpha}$. Since $(1/(2\pi)) \int V = 1$ and $\operatorname{supp} V \subseteq [-(\varepsilon/C)^{1/\alpha}, (\varepsilon/C)^{1/\alpha}]$, the smallest possible value of $V(0)$ is $\pi(C/\varepsilon)^{1/\alpha}$, i.e., $\int V = 2V(0)(\varepsilon/C)^{1/\alpha}$. By the required smoothness on V, we really have to choose V with a slightly larger value at 0 than $\pi(C/\varepsilon)^{1/\alpha}$. Thus we take any $V \in \mathscr{V}$ supported by $[-(\varepsilon/C)^{1/\alpha}, (\varepsilon/C)^{1/\alpha}]$ for which $V(0) = \pi(C/\varepsilon)^{1/\alpha} + R$, where $R = 4r\pi^2/[\varepsilon(8\pi+1)]$. Consequently, the right-hand side of (III.4.2) is $(\varepsilon(8\pi+1)/4\pi)((C/\varepsilon)^{1/\alpha} + (R/\pi))$ and we are done.

Q.E.D.

The role of part (b) is to quantify the value of λ in part (a). Also, by the discussion in Section III.2, if we are dealing with a spectral estimation problem, then S is not necessarily computable in (III.4.2) or (III.4.3) even though $S * B^2$ may be known. Theorem III.4.1 is useful in spectral estimation if, for a given problem, a suitable approximant $V \in \mathscr{V}$ to δ can be found depending, say, on the relative spacing of frequencies in the signal.

IV. Power Spectrum Computation and Wiener's Tauberian Theorem

IV.1. *A Classical Filtering Problem*

Let $x: \mathbb{R} \to \mathbb{C}$ be a sample function of the form $x(t) = y(t) + n(t)$, where y represents an intelligence-bearing real signal and n is real noise. The receiver is given x but not y or n separately. The power spectra S_x and S_y of x and y are also assumed known as well as the cross correlation

$$P_{xy}(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+u)\, y(u)\, du.$$

A classical deterministic *linear filtering problem* is to find a *filter* $f = f_m$ for which the mean-square

$$I(f) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |y(t) - f * x(t)|^2\, dt \qquad \text{(IV.1.1)}$$

is minimized, e.g., [44, Chap. 3]. There is a natural stochastic formulation of this problem where $x(t)$ is replaced by the noisy signal $x(t, \alpha) = y(t) + n(t, \alpha)$. In the stochastic case, the deterministic mean-square $I(f)$ is replaced by the probabilistic mean-square $E\{|y(t) - (f * x)(t, \cdot)|^2\}$. The solutions are formally the same for both cases, and if sufficient ergodicity is assumed, then the two models are practically equivalent, e.g., Theorem I.3.2.

In the causal situation where the filter f_m must vanish on $(-\infty, 0)$, the solution of the filtering problem is a consequence of solving a Hopf–Wiener type equation. This, in turn, can be accomplished when a certain factorization problem is solved, e.g., [31, pp. 95–97; 34, pp. 340ff.; 44, pp. 153–158 by Levinson]. The requisite factorization can be effected if the *Paley–Wiener condition*,

$$\int_{-\infty}^{\infty} \frac{\log S_x(\gamma)}{1+\gamma^2}\, d\gamma > -\infty, \tag{IV.1.2}$$

is valid, e.g., [34, Sect. 7.2]. This condition holds for rational power spectra, e.g., [18, pp. 100–108]. We mention Eq. (IV.1.2) because of its similarity to condition (III.3.4), and to point out that the precise hypotheses on S_x necessary to legitimize this filtering/factorization calculation are not known, cf. Remark IV.2.1(a). Finally, condition (IV.1.2) plays an important role in Szegö's alternative for prediction theory, e.g., [18, pp. 3–4 and Sect. 4.2].

Remark IV.1.1. In the stochastic model for the linear filtering problem, we let $x_f(t, \alpha) = (f * x)(t, \alpha)$ and consider the *bias* $b_{f,x}(t) = E\{x_f(t)\} - y(t)$ and variance $\sigma_{f,x}(t) = E\{|x_f(t) - m_{f,x}|^2\}$, where $m_{f,x} = E\{x_f(t)\} = m_x \int f$ and m_x is the mean of x. Since

$$E\{|y(t) - f * x(t)|^2\} = b_{f,x}^2(t) + \sigma_{f,x}^2(t),$$

the probabilitistic mean square, and hence its deterministic counterpart $I(f)$, is a reasonable measure of error for constructing filters.

IV.2. *The Tauberian Minimization Problem*

Let $L_+^\infty(\mathbb{R}) = \{x \in L^\infty(\mathbb{R})\colon \exists T_x$ such that $\operatorname{supp} x \subseteq [T_x, \infty)\}$ and let $L_{+*}^\infty = \{x \in L_+^\infty(\mathbb{R})\colon \exists \lim_{T\to\infty} p_T = P_x$ in the $\sigma(L^\infty(\mathbb{R}), L^1(\mathbb{R}))$ topology$\}$, where the notation p_T and $P = P_x$ was introduced in Section I.3. For consistency, we refer to P as the *autocorrelation* of x. Next, we define the functional I, introduced in (IV.1.1), as

$$I(f) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |H(t) - f * x(t)|^2\, dt \tag{IV.2.1}$$

for a fixed element $x \in L^{\infty}_{+*}$. The power spectrum S of $x \in L^{\infty}_{+*}$ exists by the discussion in Section I.3, and we shall see that the limit (IV.2.1) exists for each $f \in L^1(\mathbb{R})$. The *Tauberian minimization problem* is to characterize the criterion, $\inf\{I(f): f \in X \subseteq L^1(\mathbb{R})\} = I(g)$, $g \in X \subseteq L^1(\mathbb{R})$, in terms of the power spectrum S; and when this infimum is obtained, the filter g is said to *minimize* I in X. The solution to this problem is

THEOREM IV.2.1. *Given $x \in L^{\infty}_{+*}$ with autocorrelation P and power spectrum S, we assume the condition,*

$$M = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, dt \neq 0. \tag{IV.2.2}$$

(a) *There is a function $g \in L^1(\mathbb{R})$ which minimizes the functional I if and only if* supp $S = \{0\} \cup F$, *where $F \subseteq \hat{\mathbb{R}}$ is closed and $0 \notin F$.*

(b) *The mean M_P of the autocorrelation, defined as*

$$M_P = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} P(t)\, dt,$$

exists; *and, in the case of part* (a),

$$S = M_P \delta + \mu, \tag{IV.2.3}$$

where $\mu \in M(\hat{\mathbb{R}})$ is a positive measure supported by F and $M_P > 0$. Further, we have

$$M_P = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |f * x(t)|^2\, dt \tag{IV.2.4}$$

for all $f \in L^1(\mathbb{R})$ for which $\hat{f}(0) = 1$ and $\hat{f} = 0$ on F.

(c) *In the case of part* (a), *$M_P \geqslant 2\,|M|^2$ and $M_P = 2\,|M|^2$ if and only if $I(g) = 0$.*

(d) *A function $g \in L^1(\mathbb{R})$ is a minimizer of I if and only if $\hat{g}(0) = \bar{M}/M_P$ and $\hat{g} = 0$ on F.*

There are a number of natural questions concerning the raison d'être of the Tauberian minimization problem and the implications of its solution. We deal with this issue in the remainder of Section IV.2; the rest of Section IV is devoted to a proof of Theorem IV.2.1 (Sections IV.4–IV.6), to a method for computing power spectra (Section IV.7), and to analyzing the relation between Theorem IV.2.1, uniqueness theorems, and the distant past (Section IV.9).

Remark IV.2.1. (a) Instead of the filtering problem setup of Section IV.1, we have replaced the unknown signal $y(t)$ by the Heaviside function H in (IV.2.1). In this way, the solution of the minimization problem can be effected without restricting the power spectrum beyond the constraints of the problem itself, as opposed to external mathematical restrictions such as (IV.1.2).

(b) Let us now describe the Tauberian nature of our minimization problem. Given a sample function $x \in L^{\infty}(\mathbb{R})$, if $\lim_{t \to +\infty} x(t) = x(+\infty) \neq 0$ exists, then for all $f \in L^1(\mathbb{R})$ we have $f * x(+\infty) = x(+\infty) \int f$. If $x(+\infty)$ does not exist but $g * x(+\infty) \neq 0$ exists for some $g \in L^1(\mathbb{R})$ for which $\hat{g}$ never vanishes, then *Wiener's Tauberian theorem* asserts the existence of $f * x(+\infty)$ for each $f \in L^1(\mathbb{R})$; that is, summability of x for the kernel g implies summability of x for every integrable kernel. In both these cases, we can prove, given natural hypotheses, that the power spectrum of x is supported by the origin. If $x \in L^{\infty}_{+}$ and $g * x(+\infty)$ does not exist for any g for which $\hat{g}$ never vanishes, then the condition $I(g) = 0$ for some such g constitutes a generalization of the hypothesis to Wiener's theorem. Even more, the minimization criterion $\inf\{I(f): f \in L^1(\mathbb{R})\} = I(g)$ is a further generalization of this hypothesis. Consequently, it is natural to study the power spectrum of x when minimization by $g \in L^1(\mathbb{R})$ is possible in order to see in what way this criterion serves as an hypothesis to a generalized Wiener Tauberian theorem. To some extent, Wiener confronted a special case of this latter issue, e.g., [42; 43, p. 177].

EXAMPLE IV.2.1. The purpose of this example is to examine the role of the mean M in Theorem IV.2.1.

(a) If $M = 0$, then $g \equiv 0$ minimizes I.

(b) In order to prove Theorem IV.2.1 for the situation $\{\omega\} \cup F$ instead of $\{0\} \cup F$, it is necessary to replace (IV.2.2) by the condition $M_{\omega} = \lim(1/(2T)) \int_{-T}^{T} x(t)\, e^{-it\omega}\, dt \neq 0$. If x is uniform almost periodic, then M_{ω} exists for every $\omega \in \hat{\mathbb{R}}$ and $S = \sum M_{\omega} \delta_{\omega}$.

(c) In the case of an SSP x, it is frequently assumed that $m = E\{x\} = 0$. If $m \neq 0$, it is easy to see that the power spectrum S contains a δ. The same conclusion can be drawn if $M \neq 0$. The impact of Theorem IV.2.1 is that δ is disjoint from the remainder of S in the case of minimization.

EXAMPLE IV.2.2. (a) Given a sample function $x \in L^{\infty}_{+*}$, if $S = M_{\omega} \delta_{\omega} + \mu$, then we set $\mu = S_{\omega} = S - M_{\omega} \delta_{\omega}$. Since

$$\hat{S}(t) - M_{\omega} e^{-i\omega t} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} (x(t+u)\, \overline{x(u)} - M_{\omega} e^{-i\omega t})\, du,$$

we define x_ω by the condition, $x_\omega(t+u)\overline{x_\omega(u)}=x(t+u)\overline{x(u)}-M_\omega e^{-i\omega t}$. Thus, we obtain

$$x(t)=(M_\omega/\overline{x(0)})\,e^{-i\omega t}+(\overline{x_\omega(0)/x(0)})\,x_\omega(t). \tag{IV.2.5}$$

The power spectrum S_ω of x_ω contains no spectral lines δ_ω at ω, and $(M_\omega/\overline{x(0)})\,e^{-i\omega t}$ is a *hidden periodicity* of the sample function x.

(b) It was during the 1890s that Schuster first formulated the idea of dealing with hidden periodicities, and, even then, the natural context was in terms of an SSP x, e.g., [42, p. 127]. Thus the expression (IV.2.5) really has the form

$$x(t,\alpha)=M_\omega\phi_\omega(\alpha)+x_\omega(t,\alpha). \tag{IV.2.6}$$

(c) We continue the calculation (IV.2.6) by subtracting a spectral line δ_λ from S_ω. In this way, the *hidden periodicities* of x are represented by the *generalized almost periodic function*,

$$x_p(t,\alpha)=\sum_{\omega\in\mathfrak{P}} M_\omega\phi_\omega(\alpha)\,e^{-it\omega}.$$

Various growth hypotheses on the sequence $\{M_\omega\neq 0\}$ allow for the characterization of x_p as an element of one or the other types of almost periodic functions, cf. Section IV.6. If $x_p=x$, then the original process was essentially deterministic with no confusing noises. Theorem IV.2.1 deals with hidden periodicities whose deterministic model places them far away from the noise. Schuster's original work analyzed hidden periodicities associated with astronomical data. Besides Wiener's vital update [42, Sect. 2], there are important modern applications which depend on determining hidden periodicities, e.g., [13, pp. 510–511] and Papoulis and Chamzas (*IEEE-ASSP*, 1979).

IV.3. *Elementary Examples*

EXAMPLE IV.3.1. (a) We shall compute the power spectrum of the signal

$$x=\sum_{n=0}\chi_{[2nc,(2n+1)c)},\qquad c>0.$$

(a.i) Clearly, we see that $x^2=x$,

$$P(ck)=0 \quad \text{if } k \text{ is odd,}$$
$$\quad=\tfrac{1}{4} \quad \text{if } k \text{ is even,}$$

and P is $2c$-periodic.

(a.ii) Next we observe that P exists pointwise on $\mathbb{R}$ as a continuous function, and

$$\forall t \in [-c, c], \qquad P(t) = (1/4c)(c - |t|). \tag{IV.3.1}$$

To verify (IV.3.1), we note that if $t \in [0, c]$, then $t = ac$ for some $a \in [0, 1]$ and

$$P(ac) = \lim_{n\to\infty} \frac{1}{2(2n+1)c} \int_0^{(2n+1)c} x(u)\, x(u+ac)\, du = \frac{1-a}{4};$$

for the case $t \in [c, 2c]$, we compute $P(c + ac) = a/4$, and (IV.3.1) is obtained.

(a.iii) Because of (IV.3.1) and the periodicity of P, we have $S = (\sum_k \tau_{2kc} \Delta)\hat{\ }^{-1}$, where τ is translation and $\Delta(t) = (1/(4c))(c - |t|)\, \chi_c(t)$. Thus,

$$S = 2\pi \hat{\Delta}^{-1} \left(\sum_k \delta_{2kc} \right)\hat{\ }^{-1}. \tag{IV.3.2}$$

We compute $(\sum_k \delta_{2kc})\hat{\ }^{-1}$ and $\hat{\Delta}^{-1}$. For smooth enough f, the Poisson summation formula yields $\sum f(2kc) = (1/2c) \sum \hat{f}(\pi n/c)$ and so, by duality, we obtain $(\sum_k \delta_{2kc})\hat{\ }^{-1} = (1/(2c)) \sum \delta_{\pi k/c}$. A routine calculation gives $\hat{\Delta}^{-1}(\gamma) = (c/(8\pi))(\sin(c\gamma/2)/(c\gamma/2))^2$. Combining this information with (IV.3.2) yields the power spectrum

$$S = \frac{1}{8}\delta + \frac{1}{2\pi^2} \sum_{j=-\infty}^{\infty} \frac{1}{(2j+1)^2}\, \delta_{\pi(2j+1)c}, \tag{IV.3.3}$$

a discrete measure.

(a.iv) The above calculations give

$$\operatorname{supp} S = \{0\} \cup \{\pi(2j+1)/c : j \in \mathbb{Z}\}, \qquad M = \tfrac{1}{4} \quad \text{and} \quad M_P = \tfrac{1}{8}.$$

Thus, we have

$$M \neq 0 \qquad \text{and} \qquad \{0\} \subsetneq \operatorname{supp} S. \tag{IV.3.4}$$

Also, $I(g) = 0$ for any $g \in L^1(\mathbb{R})$ for which $\hat{g}(0) = 2$ and $\hat{g} = 0$ on $(\operatorname{supp} S)\backslash\{0\}$. In particular, there are no functions in $L^1(\mathbb{R})$ with nonvanishing Fourier transforms which minimize I.

(b) Let $x(t) = (\sin t)\, \chi_{[0,\infty)}(t)$. Then $M = 0$ and a direct calculation shows that $P(u) = r \cos u$ exists pointwise and $S = (r/2)(\delta_1 + \delta_{-1})$, where r is the power $\lim(1/(2T)) \int_0^T \sin^2 t\, dt$. In particular, $\operatorname{supp} S = \{-1, 1\}$ and

$$M = 0 \qquad \text{and} \qquad 0 \notin \operatorname{supp} S \neq \varnothing. \tag{IV.3.5}$$

(c) If $y=\sum_n(-1)^n \chi_{[n,n+1)}$, then we use H and x of part (a) to compute that the continuous function $P_y=\frac{1}{4}(P_x+1)$ exists pointwise, $S_y=\frac{1}{4}(S_x+\delta)$, and $\operatorname{supp} S_y=\{0\}\cup\{\pi(2j+1)\}$; in particular,

$$M_y=0 \qquad \text{and} \qquad \{0\}\subsetneq \operatorname{supp} S_y. \tag{IV.3.6}$$

Notice the difference between (IV.3.5) *and* (IV.3.6) *even though y is just a rectilinear version of x defined in part* (b).

(d) If $x=H$, then $P=1$ pointwise, $S=\frac{1}{2}\delta$, and $M=\frac{1}{2}$; thus,

$$M\neq 0 \qquad \text{and} \qquad \operatorname{supp} S=\{0\}. \tag{IV.3.7}$$

Since $M_p=\frac{1}{2}$, any $g\in L^1(\mathbb{R})$ for which $\hat{g}(0)=1$ is a minimizer of I.

Remark IV.3.1. The following are ancillary analytic criteria for obtaining zero value of the functional I. Details are omitted.

(a) Given $x\in L^\infty_{+*}$ and $f\in L^1(\mathbb{R})$, then $I(f)=0$ if and only if

$$\lim_{T\to\infty}\frac{1}{2T}\int_{|T|\leqslant t}(|H(t)-f*x(t)|/t)^2\,dt=0,$$

cf. [4, Exercise 2.1.2e].

(b) If $x\in L^\infty_{+*}$ and $\lim f_n*x=H$ in measure on $\mathbb{R}$, then an easy Moore–Smith argument shows that $\lim I(f_n)=0$ when $\sup\|H-f_n*x\|_\infty<\infty$. The verification of convergence in measure on *all* of $\mathbb{R}$ can be difficult.

(c) If $x=\chi_{[0,1)}$, then $H=\sum_0^\infty \delta_n*x$ and so if we approximate $\sum\delta_n$ by an element $f\in L^1(\mathbb{R})$, we would have $|\{x:|H-f*x|\geqslant\frac{1}{2}\}|=\infty$. Thus there is no convergence in measure on $\mathbb{R}$. On the other hand, if $x=\sum\chi_{[2n,2n+1)}$, then $\lim f_n*x=H$ in measure on $\mathbb{R}$ for some sequence $\{f_n\}\subseteq L^1(\mathbb{R})$ since $H=(\delta+\delta_1)*x$.

(d) Given $x\in L^\infty_{+*}$, assume M exists. If $\lim I(f_n)=0$, then $M\neq 0$.

IV.4. *Classical Criteria for Minimization*

PROPOSITION IV.4.1. *Given $x\in L^\infty_{+*}$, assume the mean M of x exists. Then for each $f\in L^1(\mathbb{R})$ we have*

$$I(f)=\tfrac{1}{2}-2\operatorname{Re}\left(M\int f\right)+\iint P(u-t)f(t)\overline{f(u)}\,dt\,du, \tag{IV.4.1}$$

where P is the autocorrelation of x. (Re$(M\int f)$ *is replaced by* $M\int$ Ref *for real-valued x.*)

Proof. (i) For $f \in L^1(\mathbb{R})$, we compute that

$$\frac{1}{2T}\int_{-T}^{T} |H(t) - f * x(t)|^2 \, dt$$

$$= \frac{1}{2} - \left(\int f(u) \left(\frac{1}{2T}\int_0^T x(t-u)\, dt\right) du \right.$$

$$+ \left.\int \overline{f(u)} \left(\frac{1}{2T}\int_0^T \overline{x(t-u)}\, dt\right) du\right)$$

$$+ \iint f(u)\overline{f(v)} \left(\frac{1}{2T}\int_{-T}^{T} x(t-u)\,\overline{x(t-v)}\, dt\right) du\, dv.$$

(ii) The result will follow from the Lebesgue dominated convergence theorem if we verify that

$$\forall u \in \mathbb{R}, \qquad M = \lim_{T\to\infty} \frac{1}{2T}\int_0^T x(t-u)\, dt. \tag{IV.4.2}$$

Since M exists, (IV.4.2) will follow once we show that $\lim(1/(2T)) \int_{-T}^{-u} x = 0$ for each fixed $u \in \mathbb{R}$. Letting $\operatorname{supp} x \subseteq [T_x, \infty]$, we have $\int_{-T}^{-u} x = 0$ if $-u \leqslant T_x$ and $|(1/(2T))\int_{-T}^{-u} x| \leqslant (-u - T_x)/(2T)$ if $-u > T_x$. This last step would not be possible without the support hypothesis on x. Q.E.D.

The proof of the next result uses Proposition IV.4.1 and a classical variational procedure, e.g., [44, Appendix C by Levinson].

PROPOSITION IV.4.2. *Given $x \in L_{+*}$, assume the mean M of x exists. Let X be a complex vector space dense in $L^1(\mathbb{R})$. Then $g \in X$ is a minimizer (in X) of $I \Leftrightarrow$*

$$\forall t \in \mathbb{R}, \qquad P * g(t) = \bar{M}, \tag{IV.4.3}$$

where P is the autocorrelation of x.

Remark IV.4.1. (a) The density of X is not required in the sufficient conditions (for minimization) of Proposition IV.4.2; and it is only used in the necessary conditions to invoke the Hahn–Banach theorem at the end of the proof.

(b) Proposition IV.4.2 is also valid when x is real valued (and therefore $\bar{M} = M$) and X is a real vector space.

(c) The proof of Proposition IV.4.2 yields the following result: Let $X = L^1[-n, \infty)$ be a complex vector space (resp., a real vector space assuming further that x is real valued) considered as a subspace of $L^1(\mathbb{R})$ by letting its elements vanish on $(-\infty, -n)$. Then $g \in X$ is a minimizer (in X) of I if and only if $P * g = \bar{M}$ (resp., M) for all $t \geqslant -n$.

IV.5. *The Solution of a Convolution Equation*

We consider the equation

$$\forall t \in \mathbb{R}, \qquad g * P(t) = M, \tag{IV.5.1}$$

where $M \neq 0$ is a given constant and $P = \hat{S}$ is a given element of $L^\infty(\mathbb{R})$. We wish to determine solutions $g \in L^1(\mathbb{R})$. In order to do this, it is convenient to introduce the following terminology: $S \in A'(\mathbb{R})$ is *synthesizable* if $\langle S, \hat{f} \rangle = 0$ whenever $\hat{f} = 0$ on supp S for $f \in L^1(\mathbb{R})$.

PROPOSITION IV.5.1. *Suppose* $S \neq \beta\delta + \mu$ *for any* $\beta \neq 0$ *and any* $\mu \in A'(\mathbb{R})$ *for which* $0 \notin \operatorname{supp} \mu$. *Then there is no* $g \in L^1(\mathbb{R})$ *for which Eq.* (IV.5.1) *is valid and* $\hat{g}(0) \neq 0$.

Proof. (i) Let $g \in L^1(\mathbb{R})$ have the property that $\hat{g}(0) \neq 0$. Then (IV.5.1) is valid if and only if

$$g * (P - (M/\hat{g}(0))) = 0 \qquad \text{on} \quad \mathbb{R}; \tag{IV.5.2}$$

and so, from the Tauberian theorem [4, Theorem 1.3.1], if (IV.5.1) is valid, then

$$\hat{g} = 0 \qquad \text{on} \quad \operatorname{supp}(S - (M/\hat{g}(0))\delta). \tag{IV.5.3}$$

(If S is synthesizable and $\hat{g}(0) \neq 0$, then (IV.5.3) is also a sufficient condition for the validity of (IV.5.1).)

(ii) We shall assume (IV.5.1) for $g \in L^1\mathbb{R}$ such that $\hat{g}(0) \neq 0$, and obtain a contradiction to the hypotheses of the proposition.

From part (i), (IV.5.1) and $\hat{g}(0) \neq 0$ imply (IV.5.3) and $\hat{g}(0) \neq 0$, and we shall obtain our contradiction assuming the conditions, (IV.5.3) and $\hat{g}(0) \neq 0$.

Note that $0 \in \operatorname{supp} S$. In fact, if $0 \notin \operatorname{supp} S$, then $\operatorname{supp}(S - (M/\hat{g}(0))\delta) = \{0\} \cup \operatorname{supp} S$, where $\{0\} \cap \operatorname{supp} S = \varnothing$; and so (IV.5.3) implies $\hat{g}(0) = 0$, a contradiction. Thus, we have the following two cases:

(iii) Assume $0 \in \operatorname{supp} S$, where 0 is a discrete point of supp S. Then $S = \beta\delta + \mu$, where $\beta \neq 0$ and $0 \notin \operatorname{supp} \mu$; but this is precisely the case we are excluding in our hypotheses and so we have the desired contradiction.

(iv) Assume $0 \in \operatorname{supp} S$, where 0 is a limit point of $\operatorname{supp} S$. Then $0 \in \operatorname{supp}(S - (M/\hat{g}(0))\delta)$ $(= \operatorname{supp} S)$, so that $\hat{g}(0) = 0$ because of (IV.5.3). This is the desired contradiction. Q.E.D.

Remark IV.5.1. (a) If $M = 0$, then $g \equiv 0$ is a solution of (IV.5.1), and, by the Tauberian theorem again, $g \equiv 0$ is the unique solution if and only if $\operatorname{supp} S = \hat{\mathbb{R}}$; in fact, there are infinitely many solutions when $\operatorname{supp} S \neq \hat{\mathbb{R}}$.

(b) Because of the condition $\hat{g}(0) \neq 0$ in Proposition IV.5.1 and the use of dense subspaces $X \subseteq L^1(\mathbb{R})$ in Proposition IV.4.2, note that if $X \subseteq L^1(\mathbb{R})$ is a dense subspace, then X contains elements g for which $\hat{g}(0) \neq 0$ (otherwise $\bar{X}$ is contained in a maximal ideal), cf. Remark IV.7.1(b).

PROPOSITION IV.5.2. *Assume* (IV.5.1) *for some g satisfying $\hat{g}(0) = 0$. Then S is not synthesizable and $S \neq \beta\delta + \mu$ for any $\beta \neq 0$ and any $\mu \in A'(\hat{\mathbb{R}})$ for which* $0 \notin \operatorname{supp} \mu$.

Proof. (i) First, suppose S is synthesizable. By hypothesis, we have $S(\hat{g})^2 = 0$, since $S\hat{g} = M\delta$ and $\hat{g}\delta = \hat{g}(0) = 0$. Thus, $\hat{g} = 0$ on $\operatorname{supp} S$ and so $S\hat{g} = 0$, since S is synthesizable. This contradicts (IV.5.1) since $M \neq 0$.

(ii) Next, assume $S = \beta\delta + \mu$ for some $\beta \neq 0$, where $0 \notin \operatorname{supp} \mu$. From (IV.5.1) and the definition of S, we have $\beta\hat{g}(0)\delta + \hat{g}\mu = M\delta$. Consequently, $\hat{g}\mu = M\delta$ since $\hat{g}(0) = 0$.

We now choose $f \in L^1(\mathbb{R})$ such that $\operatorname{supp} \hat{f} \cap \operatorname{supp} \mu = \varnothing$ and $\hat{f}(0) \neq 0$; thus $\hat{g}\hat{f}\mu = 0$. On the other hand, $\hat{g}\hat{f}\mu = M\hat{f}(0) \neq 0$ since $\hat{g}\mu = M\delta$. This is the desired contradiction and so we have proved that if $S = \beta\delta + \mu$, $\beta \neq 0$, and $0 \notin \operatorname{supp} \mu$, then there is no solution $g \in L^1(\mathbb{R})$ of (IV.5.1) for which $\hat{g}(0) = 0$. Q.E.D.

PROPOSITION IV.5.3. *If $S = \beta\delta + \mu$, where $\beta \neq 0$ and $0 \notin \operatorname{supp} \mu$, then any $g \in L^1(\mathbb{R})$ for which $\hat{g}(0) = M/\beta$ and $\operatorname{supp} \mu \cap \operatorname{supp} \hat{g} = \varnothing$ (resp., $\hat{g} = 0$ on $\operatorname{supp} \mu$ for synthesizable S) is a solution of* (IV.5.1).

Proof. Because of (IV.5.2), we must prove that $g * (P - \beta) = 0$ on $\mathbb{R}$, i.e., $\hat{g}\mu = 0$. This follows from the support hypotheses. Q.E.D.

THEOREM IV.5.1. *Assume S is synthesizable. Equation* (IV.5.1) *has a solution $g \in L^1(\mathbb{R})$ if and only if $\operatorname{supp} S = \{0\} \cup F$, where F is closed and $0 \notin F$. In this case, $S = \beta\delta + \mu$, where $\beta \neq 0$ and $0 \notin \operatorname{supp} \mu = F$; and $g \in L^1(\mathbb{R})$ is a solution of* (IV.5.1) *if and only if $\hat{g}(0) = M/\beta$ and $\hat{g} = 0$ on F.*

Proof. (i) If (IV.5.1) has a solution $g \in L^1(\mathbb{R})$, then, using the hypothesis that S is synthesizable, we see that $\hat{g}(0) \neq 0$ by

Proposition IV.5.2. Consequently, by Proposition IV.5.1, we see that $\operatorname{supp} S = \{0\} \cup F$, where F is closed and $0 \notin F$.

(ii) Conversely, if $\operatorname{supp} S = \{0\} \cup F$ then Proposition IV.5.3 shows the existence of a solution g for which $\hat{g}(0) = M/\beta$ and $\hat{g} = 0$ on F.

(iii) Finally, we must show that every solution g has the desired form. Since g is a solution, $S = \beta\delta + \mu$, where $\beta \neq 0$ and $0 \notin \operatorname{supp} \mu = F$. Suppose $\hat{g}(0) = M/\beta$ ($\neq 0$). Then (IV.5.2) and part (i) show that $\hat{g}\mu = 0$; thus, $\hat{g} = 0$ on F by the Tauberian theorem.

If $\hat{g}(0) \neq M/\beta$, we take any f for which $\hat{f}(0) \neq 0$ and $(\operatorname{supp} \hat{f}) \cap F = \varnothing$. Since g is a solution, (IV.5.2) and part (i) tell us that $\hat{g}(\mu + c\delta) = 0$, $c \neq 0$. Consequently, $\hat{f}\hat{g}(\mu + c\delta) = 0$ and so $\hat{f}(0)\, c\delta = 0$ since $(\operatorname{supp} \hat{f}) \cap F = \varnothing$; but this is the desired contradiction because f was chosen so that $\hat{f}(0) \neq 0$. Therefore we conclude that $\hat{g}(0)$ must be M/β. Q.E.D.

Remark IV.5.2. (a) The sufficient condition for a solution (in terms of support) in Theorem IV.5.1 is valid for any $S \in A'(\mathbb{R})$, cf. the comment after (IV.5.3).

(b) Summability criteria, which really reflect support properties, have been used to characterize solutions of similar but more general equations [8; 21, Theorem 4]. These results deal with arbitrary $P \in L^\infty(\mathbb{R})$ and the synthesis constraint falls on $g \in L^1(\mathbb{R})$ by means of the Beurling–Pollard integrability condition, $\int (1 + |x|)^{1/2}\, |g(x)|\, dx < \infty$.

(c) Carleman [15, pp. 115–116] gave a necessary and sufficient condition for the existence of a nonzero solution of (IV.5.2) (and therefore of (IV.5.1)) in the case that $\hat{g}(0)$ is replaced by a given nonzero constant β and $\hat{S} = P \in L^\infty(\mathbb{R})$ is arbitrary. His solution is in terms of classical analytic continuation properties of the Carleman transform defined in the upper and lower half planes. In our terms, this condition translates into the statement that $\operatorname{supp}(S - (M/\beta)\delta) \neq \mathbb{R}$. Proposition IV.5.1 is essentially a direct real-variable proof of Carleman's result.

IV.6. *The Mean of the Autocorrelation*

The following is clear from (I.3.4):

PROPOSITION IV.6.1. *Given $x \in L^\infty_{+*}$ with autocorrelation P and power spectrum S, assume $S = \beta\delta + \mu$, where $0 \notin \operatorname{supp} \mu$, and take any $f \in L^1(\mathbb{R})$ for which $\hat{f}(0) = 1$ and $\hat{f} = 0$ on $-\operatorname{supp} \mu$. Then*

$$\beta = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |f * x(t)|^2\, dt > 0.$$

PROPOSITION IV.6.2. *Given $x \in L^{\infty}_{+*}$ with autocorrelation P and power spectrum S.*

(a) *The mean M_P of the autocorrelation exists and is the mass of S at $\{0\}$.*

(b) *If $S = \beta\delta + \mu$, where $0 \notin \operatorname{supp}\mu$, then $M_P = \beta$.*

(c) *If the mean M of x exists, then*

$$M_P \geqslant |M|^2, \tag{IV.6.1}$$

and so, if $\operatorname{supp} S = \varnothing$, *then* $M = 0$, *cf.* (IV.3.4)–(IV.3.7).

Proof. (a.i) Given the Fejér kernel $f_\lambda(t) = (\lambda/2\pi)(\sin(\lambda t/2)/(\lambda t)/2)^2$, we write $\hat{f}_\lambda = \Delta_\lambda$. Then $\Delta_\lambda(\gamma) = \chi_\lambda(\gamma)(1 - |\gamma|/\lambda)$. Clearly, we have

$$\int P(t) f_\lambda(t)\, dt = (1/2\pi) \int \Delta_\lambda(\gamma)\, dS(\gamma). \tag{IV.6.2}$$

It is also easy to see, either from the definition of S and Δ_λ or using the definition of the Stieltjes integral and the fact that $F' = S$ distributionally for some increasing function F, that

$$\lim_{\lambda \to 0} \frac{1}{2\pi} \int \Delta_\lambda(\gamma)\, dS(\gamma) = S(\{0\}),$$

the mass of S at $\{0\}$. Consequently, from (IV.6.2), we obtain that

$$\lim_{\lambda \to 0} \int P(t) f_\lambda(t)\, dt = S(\{0\}). \tag{IV.6.3}$$

(a.ii) In 1926, Wiener [41] proved his direction of the Bochner–Hardy–Wiener Tauberian theorem, viz., if $P \in L^\infty[0, \infty)$ (he actually did it for the similar $P \geqslant 0$ case) and $\lim_{\lambda \to 0} \int_0^\infty f_\lambda P = A$, then

$$\lim_{T \to \infty} \frac{1}{2T} \int_0^T P = A. \tag{IV.6.4}$$

His original proof depended essentially on a Tauberian theorem of Hardy and Littlewood (1914), cf. [4, p. 91] for a proof using Wiener's general Tauberian theorem.

Since $P(t) = \overline{P(-t)}$ a.e., (IV.6.3) yields

$$S(\{0\}) = 2 \lim_{\lambda \to 0} \operatorname{Re} \int_0^\infty P(t) f_\lambda(t)\, dt. \tag{IV.6.5}$$

The existence of the limits, $\lim \int_0^\infty f_\lambda P$ (from (IV.6.3)) and $\lim(1/(2T)) \int_0^T P$ (from (IV.6.4)), yields the existence of the limits of the real parts; and so, since $(1/(2T)) \int_{-T}^T P = \mathrm{Re}(1/(2T)) \int_0^T P$, we obtain our result from (IV.6.4) and (IV.6.5). Part (b) is clear from part (a).

(c) From part (a) we know that $M_P = S(\{0\}) \geqslant 0$. Letting $y = x - M$, we see that $y \in L_*^\infty$, where there is no support condition because of M, and that $P_y = P_x - |M|^2$. By applying part (a) to y, we have $0 \leqslant S_y(\{0\}) = M_P - |M|^2$. Q.E.D.

Remark IV.6.1. If the mean M of $x \in L_{+*}^\infty$ exists and is nonzero, then we see from Proposition IV.6.2 that S has a positive mass at the origin. Of course this does not imply that $0 \notin \mathrm{supp}(S - M_P\delta)$, cf. Theorem IV.2.1(a). For example, there is a function y such that $S_y = f_1$ (the Fejér kernel), e.g., [42, pp. 203–205], and so if $x = 1 + y$, then $S_x = \delta + f_1$ and thus $0 \in \mathrm{supp}(S_x - \delta)$.

We are now able to give a complete proof of Theorem IV.2.1.

Proof (of Theorem IV.2.1). Parts (a) and (d) follow from Proposition IV.4.2, Theorem IV.5.1, and Proposition IV.6.2.

(b) Equations (IV.2.3) and (IV.2.4) follow from Propositions IV.6.2 and IV.6.1, respectively. The fact that $\mu \in M(\mathbb{A})$ is clear since $S - M_P\delta \in M(\mathbb{A})$. To show that μ is positive, let $\psi \geqslant 0$ be a function whose support contains F and excludes $\{0\}$ and which takes the value 1 on F. Then, if $\phi \geqslant 0$ is continuous, we compute $\langle \mu, \phi \rangle = \langle \mu, \phi\psi \rangle = \langle S, \phi\psi \rangle \geqslant 0$ since S is positive.

(c) Because of part (d) and Proposition IV.4.1, we see that if $g \in L^1(\mathbb{R})$ is a minimizer of I, then $I(g) = \frac{1}{2} - |M|^2/M_P \geqslant 0$. The result follows. Q.E.D.

PROPOSITION IV.6.3. *Given $x \in L_{+*}^\infty$ with autocorrelation P, assume that the mean M of x exists. If*

$$\lim_{U \to \infty} \frac{1}{2U} \int_{-U}^{U} x(t+u)\, dt = M, \qquad \textit{uniformly in} \quad u \in \mathbb{R}, \tag{IV.6.6}$$

then

$$M_P = |M|^2. \tag{IV.6.7}$$

(*Condition* (IV.6.6) *can be replaced by the weaker but more complicated condition,*

$$\lim_{U \to \infty} \frac{1}{2T} \int_0^T \left[\frac{1}{2U} \int_{-U}^{U} x(t+u)\, dt - M \right] \overline{x(u)}\, du = 0, \tag{IV.6.6$'$}$$

uniformly in $T \geqslant 1$.)

Proof. The following calculation is correct by Fubini's theorem and the Lebesgue dominated convergence theorem:

$$\begin{aligned}
|M|^2 &= \left(\lim_T \frac{1}{2T}\int_0^T \overline{x(t)}\,dt\right)\left(\lim_U \frac{1}{2U}\int_{-U}^U x(u)\,du\right)\\
&= \lim_T \frac{1}{2T}\int_0^T \overline{x(t)}\left(\lim_U \frac{1}{2U}\int_{-U}^U x(u)\,du\right)dt\\
&= \lim_T \frac{1}{2T}\int_0^T \overline{x(t)}\left(\lim_U \frac{1}{2U}\int_{-U}^U x(t+u)\,du\right)dt\\
&= \lim_T \lim_U \frac{1}{2T}\int_0^T \overline{x(t)}\left(\frac{1}{2U}\int_{-U}^U x(t+u)\,du\right)dt\\
&= \lim_T \lim_U \frac{1}{2U}\int_{-U}^U \left(\frac{1}{2T}\int_0^T \overline{x(t)}\,x(t+u)\,dt\right)du,
\end{aligned}$$

and we denote this last term by $\lim_T \lim_U a(U,T)$. Using $f=(1/(2U))\chi_{[-U,U]}\in L^1(\mathbb{R})$ and the fact that P is a weak $*$ limit, we see that for each fixed U, $\lim_T a(U,T)=(1/(2U))\int_{-U}^U P$. Also, using Fubini's theorem and the Lebesgue dominated convergence theorem, we see that for each fixed T, $\lim_U a(U,T)=M(1/(2T))\int_0^T \overline{x(t)}\,dt$, and that this convergence is uniform by (IV.6.6). Consequently, by the Moore–Smith theorem, e.g., [5, p. 236], and the above calculation, we obtain (IV.6.7),

$$|M|^2 = \lim_U \lim_T a(U,T) = \lim_U \frac{1}{2U}\int_{-U}^U P. \qquad \text{Q.E.D.}$$

Remark IV.6.2. (a) Condition (IV.6.6) is the property of uniform almost periodic functions (UAP) used to verify (IV.6.7) in the case of UAP, e.g., [42, Sect. 15].

(b) Reiter and Herz [22, Sect. 2] have used property (IV.6.6) as the basis of the notion of a maximal point spectrum sp_p, generalizing Bohr's definition for UAP. Thus, for a uniformly continuous function x, $\gamma \notin sp_p x$ if and only if $\lim(1/T)\int_u^{u+T} e^{-i\gamma t}x(t)\,dt=0$ uniformly in $u\in\mathbb{R}$; in fact, it is possible to find x for which $sp_p x=\hat{\mathbb{R}}$. On the other hand, if $x\in L^\infty(\mathbb{R})$, then $\lim(1/(2T))\int_{-T}^T e^{-i\gamma t}x(t)\,dt=0$ pointwise except possibly for small subsets of Lebesgue measure 0; such results were initiated by Herz and pursued by Eggleston and Wallin.

(c) In the case x is UAP, then P exists pointwise and is UAP. The limit $\lim(1/(2T))\int_{-T}^T x(t)\,e^{-i\gamma t}\,dt=M_\gamma$ exists for all γ and is nonzero at most countably often. Also, S is discrete and $sp_B x=\{\gamma\colon M_\gamma\neq 0\}$, the (countable) Bohr spectrum, is dense in supp S. If $x(t)=\sum_1^\infty a_n e^{it/n}$ and $\sum|a_n|<\infty$, then $x \in \mathrm{UAP}$, $a_n = M_{1/n}$, and $0 \notin sp_B x$. Thus, $M = M_0 = 0$ and $0 \in \operatorname{supp} S$, cf. Proposition IV.6.2(c).

IV.7. *Power Spectrum Computation for Summable Signals*

Given $x \in L^\infty(\mathbb{R})$, if $x(+\infty)$ exists, then $f * x(+\infty) = x(+\infty) \int f$ for each $f \in L^1(\mathbb{R})$. If $x(+\infty)$ does not exist but $g * x(+\infty) = C \int g$ exists for some $g \in L^1(\mathbb{R})$ for which $\hat{g}$ never vanishes, then $f * x(+\infty) = C \int f$ for each $f \in L^1(\mathbb{R})$. This version of Wiener's Tauberian theorem really asserts that if x is summable for certain kernels $g \in L^1(\mathbb{R})$, then it is summable for all L^1-kernels. The g-summability of $x \in L^\infty_{+*}$, represented by the existence of $g * x(+\infty) = r \neq 0$, can be generalized by the condition $I((1/r)\, g) = 0$, which, in turn, can be generalized by the condition that $(1/r)\, g$ minimizes I.

The signals $x \in L^\infty_{+*}$ for which $\operatorname{supp} S = \{0\}$ are *summable signals*. The preceding paragraph and the following theorem justify this definition, and the theorem also provides an effective means for computing S for many summable signals x. The result is really just a corollary of Theorem IV.2.1, e.g., part (d) of the proof.

THEOREM IV.7.1. *Given $x \in L^\infty_{+*}$, assume its mean $M \neq 0$ exists.*

(a) *If $x(+\infty) \neq 0$ exists, then $M = x(+\infty)/2$ and the power spectrum S equals $M_P \delta$, i.e., $|x(+\infty)|/2 = M_P$, where P is the autocorrelation of x.*

(b) *Given the characteristic function $\chi_{r/2}$, if there is a nonzero constant C such that for all small r, $\chi_{r/2} * x(+\infty) = Cr$ exists, then $|M| = |C|/2$ and $S = M_P \delta$, i.e., $|C|/2 = M_P$.*

(c) *If $g \in L^1(\mathbb{R})$ minimizes I and has a nonvanishing Fourier transform, then $S = M_P \delta$.*

(d) *If $(1/(Cr))\, \chi_{r/2}$ minimizes I for all small r, then $S = M_P \delta$, and if $I((1/(Cr))\, \chi_{r/2}) = 0$ for all small r, then $S = M_P \delta$, where $M_P = |C|^2/2$.*

Proof. (a) Given $u \in \mathbb{R}$, if we add and subtract $\overline{x(+\infty)}$ for $\overline{x(t)}$ and $x(+\infty)$ for $x(t+u)$ in $p_T(u)$ for large enough t, i.e., for $t \geqslant T_\varepsilon$ for a given ε, then it is easy to compute that $P(u) = |x(+\infty)|^2/2$.

(b.i) By our hypothesis, $I(1/(Cr))\, \chi_{r/2}) = 0$ for all small $r > 0$. Thus, by Theorem IV.2.1, $S = M_P \delta + \mu$, where $0 \notin \operatorname{supp} \mu$. In part (b(ii)) we shall show that $M_P = |C|^2/2$, so that, since $I((1/(Cr))\, \chi_{r/2}) = 0$, we have $|C|^2 = 4\,|M|^2$ by Theorem IV.2.1 again. In part (b(iii)) we shall show that $\mu = 0$.

(b.ii) Let $f = ((1/r)\, \chi_{r/2}) * f_\lambda$, where f_λ is the Fejér kernel. Since $\hat{f}_\lambda$ is the triangle $\Delta_\lambda = \chi_\lambda(\gamma)(1 - |\gamma|/\lambda)$, we have $\hat{f}(0) = 1$ and $\hat{f} = 0$ on $\operatorname{supp} \mu$ when λ is fixed small enough. For this f, Proposition IV.6.1 asserts that

$$M_P = 2\,|M|^2 = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |f * x(t)|^2 \, dt. \qquad \text{(IV.7.1)}$$

Note that $f * x(-\infty) = 0$ since $\chi_{r/2} * x(-\infty) = 0$, and, hence, by the Lebesgue dominated convergence theorem, we have

$$\lim_{t \to -\infty} ((1/r)\chi_{r/2} * x)(t-u) f_\lambda(u)\, du = 0.$$

Also, $f * x(+\infty) = C$ since $\chi_{r/2} * x(+\infty) = Cr$, and hence by the Lebesgue dominated convergence theorem,

$$\lim_{t \to \infty} ((1/r)\chi_{r/2} * x)(t-u) f_\lambda(u)\, du = Cr(1/r) \int f_\lambda = C.$$

By these properties of $f * x(\pm\infty)$ and Eq. (IV.7.1), we obtain $M_p = |C|^2/2$.

(b.iii) Because of (I.3.4) and our hypothesis,

$$\frac{(|C|\, r)^2}{2} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |\chi_{r/2} * x(t)|^2\, dt$$

$$= M_p\, |\hat{\chi}_{r/2}(0)|^2 + \int |\hat{\chi}_{r/2}(\gamma)|^2\, d\mu(\gamma)$$

for all small fixed r. Consequently, since $M_p = |C|^2/2$ and $|\hat{\chi}_{r/2}(\gamma)|^2 = (\sin^2(\gamma r/2))/(\gamma^2/4)$, we see that $|\hat{\chi}_{r/2}(0)|^2 = r^2$ and, for all small r,

$$\frac{r}{2\pi} \int \frac{\sin^2(\gamma r/2)}{(\gamma r/2)}\, d\mu(\gamma) = 0. \tag{IV.7.2}$$

By duality, this implies that $\int \hat{\mu} \Delta_r = 0$ for all small r. We showed that $\mu \geqslant 0$ and hence $\hat{\mu}(0) \geqslant 0$. If $\hat{\mu}(0) > 0$, then by the continuity of $\hat{\mu}$ we would have $\int \hat{\mu} \Delta_r > 0$ for small r. Thus, (IV.7.2) is contradicted and so $\hat{\mu}(0) = 0$. However, $\hat{\mu}$ is positive definite and so $|\hat{\mu}(t)| \leqslant \hat{\mu}(0)$ for all t. We conclude that $\mu = 0$.

(c) Since g minimizes I, we know that $0 \in \operatorname{supp} S$ from Theorem IV.2.1. Also, because of the minimization and the fact that $\hat{g}(0) \neq 0$, we know that $\hat{g}(S - (\bar{M}/\hat{g}(0))\delta) = 0$. The hypothesis that $\hat{g}$ never vanishes and the Tauberian theorem combine with this equation to imply that $\operatorname{supp}(S - (\bar{M}/\hat{g}(0))\delta) = \varnothing$. The proof of part (c) is complete.

(d) This part is a generalization of part (b) and we prove it by appealing completely to Theorem IV.2.1. In fact, μ must be zero since it cannot happen that $\hat{\chi}_{r/2} = 0$ on $\operatorname{supp} \mu$ for all small r. Q.E.D.

Remark IV.7.1. (a) Parts (a) and (b) of Theorem IV.7.1 are valid for any $x \in L^\infty(\mathbb{R})$ supported on a right half line, and it is not necessary to assume the existence of M in part (a). Also, $P = |x(+\infty)|^2/2$ pointwise in part (a). Part (b) is more interesting in this context and we can actually deduce that $x \in L^\infty_{+*}$ given that x is supported on a right half line. The proof is just a matter of doing the above proof for limit points P_α and using fixed sequences $\{T_n\}$ for which $\lim p_{T_n} = P_\alpha$ in the weak $*$ topology. In particular, each $M_{P_\alpha} = |C|^2/2$.

(b) Theorem IV.7.1(c) is really immediate from Theorem IV.2.1, and is also true when the domain $X \subseteq L^1(\mathbb{R})$ is a dense complex (resp., real if x is real valued) subspace of $L^1(\mathbb{R})$. Note that the minimizers $(1/(Cr))\chi_{r/2}$ of I which were used to prove parts (b) and (d) have Fourier transforms with zeros although, once parts (b) and (d) are proved, Theorem IV.2.1 shows there are minimizers whose Fourier transforms never vanish.

EXAMPLE IV.7.1. (a) In Section II.9, we studied the sample function

$$x = \sum_{n=0}^{\infty} \chi_{[\log(2n+1), \log 2(n+1))},$$

and its role in analytic number theory. We shall now compute its power spectrum. We first note that the continuous analogue of x is $x(t) = e^{ie^t}$. It is easy to see that $\chi_{r/2} * x(+\infty) = r/2$ exists for all $r > 0$; and, hence, from Theorem IV.7.1(b), $S = \frac{1}{8}\delta$.

(b) If $x = \sum \chi_{[2n, 2n+1)}$, then it is easy to see that $\chi_{r/2} * x(+\infty)$ does not exist for any $r \leqslant \frac{1}{4}$ and so, as we know from Example IV.3.1(a), Theorem IV.7.1(b) is not applicable. From the Tauberian theorem, we know that $f * x(+\infty)$ does not exist for any $f \in L^1(\mathbb{R})$ for which $\hat{f}$ never vanishes.

(c) It is not necessarily true that the existence of $f * x(+\infty)$ for each $f \in L^1(\mathbb{R})$ yields the existence of $x(+\infty)$, although this conclusion does follow if x is slowly oscillating, e.g., [4, Sect. 2.3]. The examples of parts (a) and (b) are not slowly oscillating.

(d) In the case where x is supported in a right half line and $f * x(+\infty) \neq 0$ exists, then $f * x$ is not a Fourier–Stieltjes transform by a classical Stone–Weierstrass and Radon–Nikodym argument.

IV.8. *The Wiener Classes and Infinite Frequencies*

For each $x \in L^2_{\mathrm{loc}}$, i.e., x is square integrable on every compact subset of $\mathbb{R}$, we consider the condition

$$\forall u \in \mathbb{R}, \qquad \exists \lim_{T \to \infty} p_T(u) = P(u). \tag{IV.8.1}$$

The *Wiener class* $\mathscr{W}$ consists of all $x \in L^2_{\mathrm{loc}}$ for which the pointwise condition (IV.8.1) holds, and $\mathscr{W}_c$ is the set $\{x \in \mathscr{W} : P \text{ is continuous}\}$. These spaces were defined and studied extensively by Wiener [42, 43, 44] and have been put in a functional analytic setting by Masani [31].

Wiener formulated an important physical distinction between the classes $\mathscr{W}_c$ and $\mathscr{W}\backslash\mathscr{W}_c$. If $x \in \mathscr{W}$, then $P(0)$ is the *total power* of the motion or signal x, and, by (I.3.4), we have, at least heuristically, that $P(0) = \int dS$. Consequently, that part of the power which lies between the frequencies $-\Omega$ and Ω is $\int_{-\Omega}^{\Omega} dS$; and analytically, $\lim_{\varepsilon \to 0} P(\varepsilon) = \lim_{\Omega \to \infty} \int_{-\Omega}^{\Omega} dS$. Thus, if $P(0) > \lim_{\varepsilon \to 0} P(\varepsilon)$, then there is a portion of the power which is not represented by any finite frequencies. In this case, signals $x \in \mathscr{W}\backslash\mathscr{W}_c$ draw part of their power from *infinite frequencies*. It is useful to compare the difference $P(0) - P(0\pm)$ with the variance $\sigma^2 = R(0) - |m|^2$ in the stochastic model.

EXAMPLE IV.8.1. (a) It is quite easy to write down elements $x \in \mathscr{W}\backslash\mathscr{W}_c$. One which Wiener used was $x(t) = e^{it^2}$, where $P = \chi_{\{0\}}$. Noting that $\int_0^\infty {\sin \atop \cos} t^2\, dt = (1/2)\sqrt{\pi/2}$, we have $M = 0$, a fact which also follows from Proposition IV.6.2(c). On the other hand, the total power of x is 1, whereas there are no associated finite frequencies, i.e., supp $S = \varnothing$.

(b.i) The logarithmic example x in Example IV.7.1(a) is not an element of $\mathscr{W}_c$. In fact, we shall see that, for each $k \geqslant 1$,

$$P(\log(2k)) = P(\log(1/2k)) = \tfrac{1}{8}, \tag{IV.8.2}$$

$$P(\log(2k+1)) = P(\log(1/(2k+1))) = (k+1)/(4(2k+1)), \tag{IV.8.3}$$

and $P(0) = \frac{1}{4}$. In particular, since $S = \frac{1}{8}\delta$, there are infinitely many points to which Wiener's *infinite frequency* model must be applied. We do not attempt a generalized model, and, of course, it may be that $x \notin \mathscr{W}$.

(b.ii) The calculations for (IV.8.2) and (IV.8.3) are technical, but the procedure is straightforward. We first write

$$\frac{1}{2T}\int_0^T x(t)\, x(t+u)\, dt = \frac{1}{2T}\int_1^{e^T} G(t)\, G(tv)\, \frac{dt}{t}, \tag{IV.8.4}$$

where $G = x \circ \log$. For each $r \in (0, 1)$, we take $u = \log r$ and so $v = r$. We also use the fact that

$$\forall c \in (0, 1), \qquad \sum_{n=1}^{N} \frac{1}{n-c} \sim \log N, \qquad N \to \infty.$$

Now if $u = \log(1/2k)$, the right-hand side of (IV.8.4) is

$$\frac{1}{2T}\sum_{n=1}^{N_T}\left[\log\frac{(4n-2)k+2}{(4n-2)k+1} + \log\frac{(4n-2)k+4}{(4n-2)k+3} + \cdots\right.$$
$$\left. + \log\frac{(4n-2)k+2k}{(4n-2)k+(2k-1)}\right]_n;$$

and since $4N_T k \leqslant e^T < 4(N_T+1)k$, we have $T \approx \log 4N_T k$. Consequently, because of the expansion of $\log(1+t)$, we have

$$\frac{1}{2\log(4Nk)}\left[\sum_{n=1}^{N}\frac{1}{(4n-2)k+1}+\sum_{n=1}^{N}\frac{1}{(4n-2)k+3}+\cdots\right.$$
$$\left.+\sum_{n=1}^{N}\frac{1}{(4n-2)k+(2k+1)}\right]_*$$
$$-\frac{1}{4\log(4Nk)}\left[\sum_{n=1}^{N}\left(\frac{1}{(4n-2)k+1}\right)^2+\cdots\right.$$
$$\left.+\sum_{n=1}^{N}\left(\frac{1}{(4n-2)k+(2k-1)}\right)^2\right]$$
$$\leqslant\frac{1}{2\log(4Nk)}\sum_{n=1}^{N}[\ ]_n\leqslant\frac{1}{2\log(4Nk)}[\ ]_*$$
$$\sim\frac{1}{8\log(4Nk)}\log N\to\tfrac{1}{8},$$

and (IV.8.2) is obtained. An analogous computation gives (IV.8.3).

PROPOSITION IV.8.1. *If x is a summable* 0–1 *signal* (*that is*, *x takes only* 0–1 *values*) *whose mean satisfies the condition $M > M_P$, then $x \notin \mathscr{W}_c$.*

Proof. From Theorem IV.2.1, $M_P = \|S\| = \hat{S}(0)$ and since x is a 0–1 signal, we have $M = P(0)$. By the hypothesis $M_P < M$, we have $P(0) > \hat{S}(0)$ and thus $x \notin \mathscr{W}_c$. Q.E.D.

Thus any 0–1 signal satisfying the conditions of Theorem IV.7.1(b),(d) is not an element of $\mathscr{W}_c$.

Remark IV.8.1. Bass [3] considers those elements $x \in \mathscr{W}_c$ whose mean M exists. He uses such functions in his model of turbulent behavior. In fact, if $x \in \mathscr{W}_c$ has autocorrelation P and $P(+\infty) = |M|^2$, then he supposes that x represents a velocity component of a turbulent fluid and constructs turbulent solutions of the Navier–Stokes equations by means of these functions.

IV.9. *The Role of the Distant Past for Tauberian Minimization*

The following result, which is naturally formulated in terms of the Tauberian minimization problem, depends on results that can be viewed, with an eye toward generalization, as elementary sampling theorems, cf. the discussion in Section III.3.

THEOREM IV.9.1. *Given* $x \in L^{\infty}_{+*}$, *let* $\operatorname{supp} S = \{0\} \cup F$, *where* F *is closed and* $0 \notin F$. *Let* $L^1_+ = \{f \in L^1(\mathbb{R})\colon \exists T_f$ *such that* $\operatorname{supp} f \subseteq [T_f, \infty)\}$.

(a) *If* I *is minimized by* $g \in L^1_+ \backslash \{0\}$, *then* F *is totally disconnected.*

(b) *If* I *is minimized by an element* $g \in L^1_+ \backslash \{0\}$ *for which* $\hat{g} \in L^1(\mathbb{R})$, *then* $|F| = 0$, *where* $|F|$ *denotes the Lebesgue measure of* F. ($|F| = 0$ *implies* F *totally disconnected, but not vice versa.*)

Proof. (a.i) Since $g \in L^1_+$, then $\hat{g}$ vanishes on at most a closed totally disconnected set. This is a classical fact which we discuss in part (a.ii). By Theorem IV.2.1, $\hat{g} = 0$ on F and so the result follows.

(a.ii) There are several ways of verifying the fact about totally disconnected zero-sets in part (a.i). We choose a sledge hammer due to Beurling [8], since it is possible to use Beurling's results and extensions of it for generalizations of Theorem IV.9.1, cf. [7] and Section III.3.

(a.iii) A version of Beurling's theorem is the following: Let S be a tempered distribution and suppose that $(\operatorname{supp} S)^{\sim}$ contains the set $\bigcup_{n=1}^{\infty} [c_n - r_n, c_n + r_n]$, where $0 < c_1 < c_2 < \cdots$ and $\sum (r_n/c_n)^2 = \infty$; then $S = 0$ if $\hat{S} = 0$ on some open interval.

Since the minimizer $g \in L^1_+ \backslash \{0\}$ has the property that $\hat{g}$ vanishes on F, we shall suppose $\hat{g}$ vanishes on some interval and prove that $g \equiv 0$, a contradiction.

Suppose $(\operatorname{supp} g)^{\sim} \supseteq (-\infty, -N]$ since $g \in L^1_+$. Then

$$\bigcup_n [c_n - r_n, c_n + r_n] \subseteq (-\infty, -N],$$

where $c_n = -N - n - 1$ and $r_n = \sqrt{n}$; hence $\sum (r_n/c_n)^2 = \infty$. Thus, Beurling's theorem applies and so $g \equiv 0$.

(b) By Theorem IV.2.1, we know that $\hat{g} = 0$ on F. If $|F| > 0$, then the second theorem of F. and M. Riesz (Remark II.1.2(b)) and the support and integrability hypotheses on g and $\hat{g}$ allow us to conclude that $g = 0$ a.e., a contradiction. Q.E.D.

Remark IV.9.1. (a) The hypothesis $g \in L^1_+$ reflects the suggestive language that g contains no information from the distant past. Thus, information from the distant past is required to minimize I for signals whose power is generated from large (thick) sets of frequencies.

(b) In [6], we have noted the important role of the distant past in closure problems associated with weighted L^1.

Acknowledgment

I gratefully acknowledge valuable insights from the following colleagues at the University of Maryland and at MIT dealing with various issues related to this paper, especially the implementation of my algorithm: C. Braccini, M. Brachet, B. Cipra, J. Evans, W. Evans, A. Fowler, J. McClellan, J. McLaughlin, A. V. Oppenheim (and his Digital Signal Processing seminar), and J. Shearer.

References

1. A. Arcese, On the method of maximum entropy spectrum estimation, *IEEE Trans. Inform. Theory*, to appear, 1983.
2. A. Baggeroer, Confidence intervals for regression (MEM) spectral estimates, *IEEE Trans. Inform. Theory* **IT22** (1976), 534–545.
3. J. Bass, Stationary functions and their applications to turbulence II, *J. Math. Anal. Appl.* **47** (1974), 458–503.
4. J. Benedetto, "Spectral Synthesis," Academic Press, New York, 1975.
5. J. Benedetto, "Real Variable and Integration," Teubner, Stuttgart, 1976.
6. J. Benedetto, A closure problem for signals in semigroup invariant systems, *SIAM J. Math. Anal.* **13** (1982), 180–207.
7. M. Benedicks, "Supports of Fourier Transform Pairs and Related Problems on Positive Harmonic Functions," TRITA-MAT-1980-10, Royal Institute of Technology, Stockholm, 1980.
8. A. Beurling, Sur la composition d'une fonction sommable et d'une fonction bornée, *C. R. Acad. Sci. Paris* **225** (1947), 274–275.
9. A. Beurling, "On Quasianalyticity and General Distributions," Summer Institute, Stanford University, 1961.
10. T. Bishop and T. Ulrych, Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space Phys.* **13** (1975), 183–200.
11. R. Blackman and J. W. Tukey, "The Measurement of Power Spectra," Dover, New York, 1959.
12. L. deBranges, "Hilbert Spaces of Entire Functions," Prentice–Hall, Englewood Cliffs, N.J., 1968.
13. D. Brillinger, "Time Series, expanded ed.," Holden–Day, Inc., San Francisco, 1981.
14. P. Buchhave, W. George, and J. Lumley, The measurement of turbulence with the laser-Doppler anemometer, *Ann. Rev. Fluid Mech.* **11** (1979), 443–503.
15. T. Carleman, "L'intégrale de Fourier (1935)," Inst. Mittag–Leffler, 1967.
16. D. Childers, Ed., "Modern Spectrum Analysis," IEEE, New York, 1978.
17. J. Doob, Wiener's work in probability theory, *Bull. Amer. Math. Soc.* **72** (1) (1966), 69–72.
18. H. Dym and H. McKean, "Gaussian Processes, Function Theory, and the Inverse Spectral Problem," Academic Press, New York, 1976.
19. P. Fougere, A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis, *J. Geophys. Res.* **82** (1977), 1051–1054.
20. M. Gaster and J. B. Roberts, On the estimation of spectra from randomly sampled signals: a method of reducing variability, *Proc. Roy. Soc. London* **371** (1980), 235–258.
21. C. Herz, A note on summability methods and spectral analysis, *Trans. Amer. Math. Soc.* **86** (1957), 506–510.
22. C. Herz, The spectral theory of bounded functions, *Trans. Amer. Math. Soc.* **94** (1960), 181–232.

23. K. Hoffman, "Banach Spaces of Analytic Functions," Prentice–Hall, Englewood Cliffs, N.J., 1962.
24. P. Koosis, Harmonic estimation in certain slit regions and a theorem of Beurling and Malliavin, *Acta Math.* **142** (1979), 275–304.
25. P. Krishnaiah (Ed.), "Developments in Statistics," Vol. 1, Academic Press, New York, 1978.
26. R. Lacoss, "Data Adaptive Spectral Analysis Methods," *Geophys.* **36** (1971), 661–675.
27. J. Lamperti, "Stochastic Processes," Springer-Verlag, New York, 1977.
28. N. Levinson, "Gap and Density Theorems," *Amer. Math. Soc. Colloq. Publ.* **26** (1940).
29. G. W. Mackey, Harmonic analysis as the exploitation of symmetry—a historical survey, *Bull. Amer. Math. Soc.* **3** (1980), 543–698.
30. P. Masani, Wiener's contribution to generalized harmonic analysis, prediction theory, and filter theory, *Bull. Amer. Math. Soc.* **72** (1) (1966), 73–125.
31. P. Masani, Commentary on the memoire [30a] on generalized harmonic analysis, *in* "Norbert Wiener: Collected Works II," pp. 333–339, MIT Press, 1979.
32. A. V. Oppenheim and R. Schafer, "Digital Signal Processing," Prentice–Hall, Englewood Cliffs, N.J., 1975.
33. A. V. Oppenheim, Ed., "Applications of Digital Signal Processing," Prentice–Hall, Englewood Cliffs, N.J., 1978.
34. A. Papoulis, "Signal Analysis," McGraw–Hill, New York, 1977.
35. H. Pollard, The Bernstein approximation problem, *Proc. Amer. Math. Soc.* **6** (1955), 402–411.
36. R. Redheffer, Completeness of sets of complex exponentials, *Advan. Math.* **24** (1977), 1–62.
37. L. Schwartz, "Théorie des Distributions," Hermann, Paris, 1966.
38. F. Tseng, T. Sarkar, and D. Weiner, A novel window for harmonic analysis, *IEEE Trans. Acoust., Speech, Signal Process.* **29** (1981), 177–188.
39. J. W. Tukey, An introduction to the measurement of spectra, *in* "The Cramér Volume" (U. Grenander, Ed.), pp. 300–330, Wiley, New York, 1959.
40. J. W. Tukey, An introduction to the calculation of numerical spectrum analysis, *in* "Spectral Analysis of Time Series" (Bernard Harris, Ed.), Wiley, New York, 1966, 25–46.
41. N. Wiener, On a theorem of Bochner and Hardy, *J. London Math. Soc.* **2** (1927), 118–123.
42. N. Wiener, Generalized harmonic analysis, *Acta Math.* **55** (1930), 117–258.
43. N. Wiener, "The Fourier Integral and Certain of its Applications," Dover, New York, 1933.
44. N. Wiener, "Time Series (1949)," MIT Press, Cambridge, 1964.